

Montpellier SupAgro
Centre international d'études supérieures en sciences agronomiques

THÈSE

Pour obtenir le grade de

DOCTEUR DE MONTPELLIER SUPAGRO

Ecole Doctorale : SIBAGHE (Systèmes Intégrés en Biologie, Agronomie, Géosciences,
Hydrosciences, Environnement)

Spécialité : EERGP (Evolution, Ecologie, Ressources génétiques, Paléontologie)

Préparée au sein de l'UMR Biologie et Génétique des Interactions Plantes-Parasites

Présentée et soutenue publiquement le: 17 juin 2011

Par :

Adrien RIEUX

Etude des processus de dispersion et des flux géniques chez un
champignon phytopathogène: le cas de *Mycosphaerella fijiensis* à
l'échelle d'un bassin de production Camerounais.

Devant le jury composé de :

Dr François BALLOUX, Imperial College, Londres – RAPPORTEUR
Dr Thomas GUILLEMAUD, INRA, Sophia Antipolis – RAPPORTEUR
Dr François DELMOTTE, INRA, Bordeaux – EXAMINATEUR
Dr Jacques DAVID, INRA, Montpellier – EXAMINATEUR
Dr Jean CARLIER, CIRAD, Montpellier – DIRECTEUR DE THESE
Dr Virginie RAVIGNE, CIRAD, Montpellier – CO-ENCADRANTE DE THESE

Membres invités au Jury :

Dr Luc DE LAPEYRE DE BELLAIRE, CIRAD, Montpellier – CO-ENCADRANT DE THESE
Dr Andreas MEHL, Bayer CropScience, Monheim (Allemagne) – PARTENAIRE de la bourse CIFRE

Montpellier SupAgro
Centre international d'études supérieures en sciences agronomiques

THÈSE

Pour obtenir le grade de

DOCTEUR DE MONTPELLIER SUPAGRO

Ecole Doctorale : SIBAGHE (Systèmes Intégrés en Biologie, Agronomie, Géosciences,
Hydrosciences, Environnement)

Spécialité : EERGP (Evolution, Ecologie, Ressources génétiques, Paléontologie)

Préparée au sein de l'UMR Biologie et Génétique des Interactions Plantes-Parasites

Présentée et soutenue publiquement le: 17 juin 2011

Par :

Adrien RIEUX

Etude des processus de dispersion et des flux géniques chez un champignon phytopathogène: le cas de *Mycosphaerella fijiensis* à l'échelle d'un bassin de production Camerounais.

Devant le jury composé de :

Dr François BALLOUX, Imperial College, Londres – RAPPORTEUR
Dr Thomas GUILLEMAUD, INRA, Sophia Antipolis – RAPPORTEUR
Dr François DELMOTTE, INRA, Bordeaux – EXAMINATEUR
Dr Jacques DAVID, INRA, Montpellier – EXAMINATEUR
Dr Jean CARLIER, CIRAD, Montpellier – DIRECTEUR DE THESE
Dr Virginie RAVIGNE, CIRAD, Montpellier – CO-ENCADRANTE DE THESE

Membres invités au Jury :

Dr Luc DE LAPEYRE DE BELLAIRE, CIRAD, Montpellier – CO-ENCADRANT DE THESE
Dr Andreas MEHL, Bayer CropScience, Monheim (Allemagne) – PARTENAIRE de la bourse CIFRE

Résumé

La dispersion est un processus clef dans la dynamique et l'évolution des populations naturelles. En plus de son rôle primordial dans les processus de colonisation, la dispersion influence également les processus d'adaptation des organismes. Chez les pathogènes, une meilleure compréhension des processus de dispersion apparaît de ce fait être un enjeu majeur pour mieux les contrôler. Durant cette thèse, nous avons étudié les processus de dispersion et quantifié les flux de gènes qui en découlent chez le champignon parasite du bananier *Mycosphaerella fijiensis*. Cette étude a été réalisée à l'échelle locale d'un bassin de production du Cameroun (la région dite du Moungo) et nous avons combiné plusieurs approches complémentaires considérant différentes échelles spatio-temporelles. Dans un premier temps, nous avons décrit, à l'aide de marqueurs génétiques neutres, la structuration spatiale des populations de *M. fijiensis* dans la région du Moungo qui présente différentes barrières potentielles à la dispersion. Nous n'avons décelé aucun effet du paysage ni de la distance géographique sur la structuration génétique. Cependant, une rupture spatiale dans les fréquences alléliques, vraisemblablement de nature historique a été mise en évidence. Ces résultats suggèrent l'existence de grandes populations de *M. fijiensis* s'écartant de l'équilibre mutation-dérive. Dans un second temps, nous avons utilisé la théorie des clines génétiques pour étudier les forces à l'origine de la mise en place et de l'évolution de gradients spatiaux de fréquences alléliques. En particulier, l'analyse de la variation spatio-temporelle de la discontinuité génétique précédemment détectée par un modèle de clines neutres nous a permis d'estimer l'intensité des flux géniques ($\sigma = 1175$ m/génération). Finalement, nous avons mesuré la distribution des distances de dispersion des deux types de spores produites par *M. fijiensis* à partir d'une source d'inoculum primaire. Cette expérimentation nous a permis de confirmer que les ascospores participent à une dispersion à grande distance alors que les conidies sont impliquées dans une dispersion à très courte distance. Nous avons estimé une distance moyenne de dispersion de 3,12 et de 283 mètres/génération respectivement pour les conidies et les ascospores et montré que le noyau de dispersion des ascospores est caractérisé par une queue lourde. Cette thèse a permis de préciser comment *M. fijiensis* se disperse et les estimations réalisées pourront être intégrées dans des modèles théoriques afin de mieux comprendre l'évolution des résistances aux fongicides et de définir des stratégies durables d'utilisation raisonnée des traitements chimiques.

Mots clefs

Dispersion, colonisation, flux de gènes, structure génétique spatiale, clines, noyau de dispersion, résistance aux fongicides, *Mycosphaerella fijiensis*, bananier.

Summary

Dispersal is a key process for both the dynamics and evolution of natural populations. In addition to being crucial for colonization, dispersal also influences the processes occurring during adaptation. For pathogens, a better understanding of dispersal processes may improve our capacity to control the diseases that they cause. In this thesis, we studied dispersal processes and quantified gene flow in the banana plant pathogen *Mycosphaerella fijiensis* at the local scale of a production area in South-West Cameroon (named Moungo). For this purpose, several approaches differing in the spatio-temporal scale to which they refer were combined. First, neutral markers were used to describe the spatial genetic structure of this pathogen in the Moungo area, which includes several potential ecological barriers to dispersal. No effects on genetic structure of landscape elements or geographical distance were found. However, we detected a spatial break in allelic frequencies that appeared to be explained by an historical event. This result suggests the existence of large *M. fijiensis* populations out of the mutation-migration-drift genetic equilibrium. Second, genetic cline theory was applied to study the evolutionary forces implicated in the installation and evolution of spatial gradients in allelic frequencies. More specifically, we analysed the spatio-temporal variation of the genetic discontinuity previously detected through a neutral cline model to estimate the intensity of gene flow in this area ($\sigma = 1175$ m/generation). Lastly, we measured the distribution of dispersal distances of *M. fijiensis* spores from a primary source of inoculum was. Such an experiment allowed us to confirm that conidia are implicated in short-distance dispersal whereas ascospores are responsible for spread of the disease over longer distances. The estimated mean dispersal distance travelled by spores was 3.12 and 283 metres/generation for conidia and ascospores, respectively, and the ascospore dispersal kernel was shown to be fat-tailed. This thesis adds to global knowledge of *M. fijiensis* dispersal and the measures of dispersal estimated in this work will be useful in parameterizing models aimed at a better understanding of the spatial patterns of fungicide resistance evolution under different management strategies.

Key words

Dispersal, colonization, gene flow, spatial genetic structure, clines, kernel of dispersal, fungicide resistance, *Mycosphaerella fijiensis*, banana plant.

MERCI,

Certains pourraient penser qu'avoir 3 encadrants durant une thèse peut être un fardeau (3 fois plus d'avis et de corrections parfois contradictoires pendant la rédaction par exemple !). Pour ma part, ce fut un réel plaisir de vous avoir eu tous les 3 à mes côtés pendant ces 1095 jours : Jean, Virginie et Luc : MERCI pour l'excellent encadrement que vous m'avez dispensé. Jean, merci pour avoir toujours répondu présent, pour m'avoir fait confiance et pour tes conseils. Je retiendrai sans aucun doute ton désormais fameux "schéma obsessionnel" ainsi que nos descentes sur les pistes d'Aussois (tu as encore un peu d'entraînement à faire pour pouvoir me rattraper, mais j'avoue que tu es sur la bonne voie...). Je t'ai probablement fait rire (ou pas) quelquefois (le coup de l'assurance avant la mission au Cameroun, ou encore lors de mon second comité sans salle ni vidéoprojecteur réservé) mais tu m'as battu à plate couture en ratant ton avion pour les States, même si je persiste à penser qu'il s'agit d'un acte manqué pour ne pas abandonner ton étudiant en fin de rédaction de thèse ! Je viens de clôturer ton compte « chewing-gums »: le crédit (à solder au plus vite) est de 272 unités ! Virginie, merci pour ta patience, ta disponibilité et ton appui. Ta culture encyclopédique de fervente évolutionniste m'a souvent inspiré et ton regard de théoricienne a permis de donner une autre dimension à cette thèse. D'un point de vue plus pratique, ton "super-pouvoir" consistant à solutionner de longues discussions en quelques « clics & scripts » sous Mathematica m'a été très utile. J'avoue que tu as, à plusieurs reprises accompli l'exploit de me faire aimer les maths...et ça c'était pas gagné d'avance ! Je te souhaite beaucoup de bonnes choses pour la suite. Luc, tu es pour beaucoup dans le montage du projet ATF à l'origine de cette thèse. Merci pour m'avoir donné ma chance. Merci également pour m'avoir fait découvrir ce merveilleux pays qu'est le Cameroun. Je garde un excellent souvenir des moments que l'on a passés ensemble sur le terrain (incluant les quelques nuits passées à découper des lésions !). Tu es et resteras pour moi "L'Indiana Jones de la Banane !"...

Je tiens ensuite à remercier l'entreprise Bayer CropScience pour m'avoir financé pendant ces trois années de thèse. Andreas and Klaus, many thanks for your financial and technical support. Thanks also for the pubs-tour trip you guys organized in Monheim. Jean and I keep an excellent souvenir of this time (and I also keep an excellent souvenir of Jean during this night!). I also want to thanks Martin and all the staff member of the lab for the pyrosequencing work. I hope the results obtained in this thesis will be helpful in a close future.

D'autres chercheurs ont également fortement contribué au travail réalisé durant cette thèse. Tout d'abord, Fabien, merci pour m'avoir donné goût à la génétique des populations spatiale lors de mon stage de DEA et également pour avoir répondu présent à chaque fois que je t'ai sollicité durant cette thèse. Ce premier papier aura mis du temps à prendre forme, mais je suis fier d'en être venu à bout avec toi (et j'espère que l'on remettra ça !). Je remercie également les chercheurs qui ont accepté de participer à mes 3 comités de thèses: Anne-Sophie, Denis, Etienne et Thomas. Merci car les discussions que nous avons eues ensemble ainsi que les points de vue et les conseils que vous m'avez donnés ont toujours été bénéfiques. Thomas merci pour ta disponibilité. C'est en grande partie grâce à toi que j'ai

commencé à comprendre la méthode du maximum de vraisemblance et pris par la suite beaucoup de plaisir à la manipuler dans le cadre de la théorie des clines génétiques (mon coup de coeur de la thèse !). Finalement, je remercie vivement les chercheurs qui ne m'ont jamais fermé leur porte (bien au contraire) et qui ont répondu à mes innombrables questions: Renaud et Arnaud du CBGP, François, Gaël, David et Philippe de BGPI, François Rousset et Samuel d'Avignon: Merci pour votre patience et votre aide. Tous nos échanges ont grandement influencé mon apprentissage.

J'adresse un grand merci au personnel de l'UMR BGPI et en particulier aux membres de l'équipe Becq pour leur accueil chaleureux. Marie, merci pour ton soutien technique et bravo pour t'être plongée dans les analyses en génétique des populations. Je suis impressionné par la vitesse à laquelle tu as récemment acquis toutes ces compétences (attention n'en fais pas trop, laisse un peu de boulot aux thésards !). La quantité de chewing-gums que Jean doit me reverser ne suffira probablement pas à solder mon crédit chez toi !...Je te souhaite une très bonne continuation. Rémos un grand merci pour ton aide dans la réalisation de la manip de suivi direct de la dispersion au Cameroun. Les bananiers artificiels que tu nous as confectionnés ont participé pour beaucoup au succès de cette expérience. Je garderai d'excellents souvenirs des quelques jours que nous avons passés là-bas ensemble. D'un point de vue strictement moins professionnel merci pour ton aide dans l'entretien mécanique de ma petite voiture et de ma vieille Vespa. Merci également à Luc (Allez Paillade !), Riton (Le smasheur fou), Véro et Anny. Je remercie aussi les chercheurs de l'équipe : Catherine, Françoise, Didier, Babeth, Daniel et Michel (proximalement vôtre!) pour m'avoir supporté et soutenu. Un immense MERCI à Geneviève mais également à Flo, Domi et Marie-Carmen qui d'un point de vue administratif et logistique font de l'UMR BGPI un labo formidable pour réaliser une thèse. Finalement, merci aux copains du labo et aux compagnons de bureaux : le trio infernal Ricou - Jean Phi_(on) - Benoît pour les pauses-café (olé !), -verveine et -ping-pong, Flo (pour tes histoires incroyables avec la badgeuse...on pourrait en faire un film), Imène (entre autre pour avoir donné naissance à la désormais célèbre attaque COBRA), Ju, Xavier, Steph, Amandine, Pierre-o, Guy, et Virginie. Pour en finir avec BGPI, je garderai d'excellents souvenirs des moments passés autour de sa table de ping-pong, notamment grâce à Benoit et Ricou (les inventeurs de l'échelle dite de la « fessée»), JPP (à l'origine de la « Jean-Yon ») et Flo (pour une certaine histoire de couches-culottes)

Cette thèse a également donné naissance à une formidable rencontre avec l'Afrique au travers des 4 missions que j'ai eu la chance de réaliser au Cameroun. Un immense Merci à Josué, mon acolyte, qui m'a accompagné dans les différentes campagnes d'échantillonnages et fait découvrir quelques-unes des nombreuses richesses de son pays. Djo, bravo pour le boulot que tu réalises au CARBAP et bon courage pour la suite de la thèse. J'adresse aussi un grand merci au personnel de l'équipe phyto du CARBAP : Oscar, Flo, Thomson, Joseph, Sakio, Robert, Henry, Clément, Lévi, Adolphe et Tade sans qui la réalisation des différents échantillonnages et de l'expérimentation directe n'aurait très certainement pas été envisageable. Merci pour votre gentillesse, votre aide et votre hospitalité. Je n'oublierai jamais les trois cérémonies de la MAYA auxquelles j'ai eu la chance de participer. Merci également aux deux expats CIRAD, Sébastien et Sylvain pour leur aide logistique sur

place. Finalement, un clin d'oeil à Clémence, Georges et Irène pour tous les moments inoubliables que j'ai passés grâce à vous en terre camerounaise.

Et puis il y a ceux sans qui cette thèse aurait été nettement moins agréable...LES AMIS. Tout d'abord merci à l'équipe avignonnaise, malgré notre éloignement géographique, quel bonheur à chacune de nos retrouvailles (cuir_moustache) : Ben & O2 (2011 restera pour sûr une grande année), Peko (big up à notre combo d'arnaques pendant ces vacances en Guadeloupe), Bouti, Rhô & Fî, Dibi & Cina, B'anus, Thib & ash (merci pour les accueils à London), Nono (ksss) et So la marmotte, LATZ (le phoenix). Ensuite il y a un sacré nombre de Montpelliérains qu'il m'est impossible de ne pas remercier. Ève pour commencer, que de bons moments passés depuis nos premiers fous rires sur les bancs de la fac. Je suis certain qu'en cumulant toutes nos soirées au fizz on remplit au moins deux disquettes (toi seule me comprendra ici je pense, et cela me fait bien plaisir). Il me faudrait faire un chapitre entier pour pouvoir énoncer tous ces souvenirs et je sais également que je devrais laisser de nombreuses pages blanches car notre histoire commence juste ! Bon courage pour ta fin de thèse. Ricou ensuite, inutile de dire qu'il faudrait être sacrément tordu pour t'inventer si tu n'existais pas ! Tu es et resteras mon idole ! YMMD, trop dur de choisir parmi toutes tes ricoulades mais bon... (il y a le cobra fasciste ! et puis deux ordis quand même...mince ! j'attends le troisième - olé). Merci également à la fine équipe du CBGP (le moins qu'on puisse dire c'est qu'il y a de l'ambiance là-bas !) : Marie (on retiendra notre écrasante victoire au dîner presque parfait, et merci pour ton soutien culinaire pendant ma phase de rédaction), Ju (bientôt de retour...cool !), Bébèr (mais quelle arnaque...), Gaël, Fab & Olive (bientôt vot' tour ! courage), Manu, Asraff, Max (118-218...toutouyoutou), Anne (les deux), Taï & Gauthier, ALC, Caro, Philippe, Laurent, Yannick, Mika, Armelle, Benoit, Réjane, Nath... Que de bons moments passés durant toutes ces bières et pétanques parties ! Merci également à ceux du côté du CEFÉ : Ben (entre beaucoup d'autres choses, merci pour avoir relu mon intro beuddy...YGM), Delph & Thom, Cat. Merci à Angelo (Forza pour cette fin de thèse...) et puis à l'équipe des beach volleyeurs (Pauline, Nico, Cyrielle, Alex, Justin, Vlad, Pierrick, Justine, Jérôme, Bastos et les autres). Tusen Takk à Sara (que de souvenirs lors de ces vacances en Norvège !), Alice (ça glisse au pays des bouteilles), et les deux champions Guigui et Captain' Raph. Merci également à tous mes collocs, Dav (bien plus qu'un colloc !), Petra, Beuddy, Kim et Kjersti pour votre soutien.

Cette thèse marque la fin de ma vie d'étudiant. Je ne peux donc pas oublier de remercier ma famille, pour sa présence et son soutien tout au long de ces nombreuses années d'études. Merci à mes parents, à Lucille, à Mone, Faby, Milou et Jackie pour m'avoir transmis une des clefs de la réussite : rien ne s'obtient sans travail.

Je terminerai en remerciant François Balloux et Thomas Guillemaud de bien vouloir être rapporteurs de cette thèse ainsi que François Delmotte et Jacques David de faire partie du jury.

Finalement, j'ose penser que les personnes que j'ai malencontreusement oubliées me pardonneront... J'en profite pour les remercier !

SOMMAIRE

<u>Contexte général</u>	9
<u>Introduction générale</u>	12
<u>I- Dispersion et flux géniques</u>	12
<u>I-1 - Quelques définitions</u>	12
<u>I-2 - La dispersion au cœur de la génétique des populations</u>	12
<i>I-2-a : Dispersion et colonisation</i>	14
<i>I-2-b : Dispersion et flux géniques intra et inter populations</i>	16
<u>I-3 - Mesurer la dispersion</u>	16
<i>I-3-a : Quels paramètres de dispersion peut-on estimer ?</i>	17
<i>I-3-b : Les méthodes directes</i>	19
<i>I-3-c : Les méthodes indirectes</i>	19
<i>I-3-d : Synthèse des différences entre approches directes et indirectes</i>	24
<u>II- Structure génétique spatiale des populations dans la réalité des paysages</u>	26
<u>III- Le modèle d'étude: Un champignon phytopathogène aérien dans son paysage agricole</u>	28
<u>III-1- L'étude de la dispersion chez les champignons phytopathogènes</u>	28
<u>III-2- Le pathosystème : <i>Mycosphaerella fijiensis</i> / bananier</u>	30
<u>III-3- Dispersion et flux de gènes chez <i>M. fijiensis</i> : Etudes antérieures</u>	33
<u>III-4- La zone de production Camerounaise du Mounjo (site d'étude)</u>	35
<u>IV- Objectifs de la thèse</u>	37

<u>Chapitre 1 : Etude de la structure génétique spatiale des populations</u>	39
Présentation générale.....	40
<u>Article 1</u> : Inferences on pathogenic fungus population structures from microsatellite data: new insights from spatial genetics approaches.....	43
<u>Article2</u> : No effect of recent range expansion and fungicide treatment on the spatial genetic structure of a plant pathogenic fungus in its agricultural landscape.....	57
Document annexe 1.....	72
Document annexe 2.....	78
<u>Chapitre 2 : Inférence de paramètres à partir des modèles de clines génétiques</u> ...	81
Présentation générale.....	82
<u>Article 3</u> : On the use of temporal variation in neutral genetic clines to estimate gene flow: a case study in a fungal plant pathogen.....	85
<u>Article 4</u> : Spatio temporal distribution and evolution of fungicides resistances alleles at the interface between treated and untreated populations of a plant pathogenic fungus.....	109
<u>Chapitre 3 : Estimation directe d'un noyau de dispersion</u>	131
Présentation générale.....	132
<u>Article 5</u> : Spore dispersal in a plant pathogenic fungus: estimation of anisotropy and individual dispersal functions from one-generation disease gradients.....	134
<u>Discussion générale</u>	163
<u>Références bibliographiques</u>	175
<u>Document annexe 3</u> (Article 6)	185

Contexte général

Les maladies de plantes sont à l'origine de nombreuses pertes en agriculture (se chiffrant en millions de tonnes de pertes annuelles) et les champignons sont responsables de près de la moitié des pathologies (Lepoivre 2003, Anderson *et al.* 2004). Bien que les exemples de champignons phytopathogènes ne manquent pas (Tableau 1), deux événements historiques sont souvent cités pour illustrer les conséquences dramatiques qu'ils peuvent potentiellement engendrer. Le premier est lié à l'altération de la biodiversité lorsque l'ascomycète *Cryphonectria parasitica* a causé la disparition quasi totale du châtaigner en Amérique du nord à la fin du XIX^{ème} siècle (Milgroom 1996). Le second illustre les risques associés à la production de denrées alimentaires de base lorsque l'oomycète *Phytophthora infestans*, agent causal du Mildiou de la pomme de terre a engendré la grande famine du début du XIX^{ème} siècle en Irlande (Birch *et al.* 2001). Face à ces menaces, la phytopathologie a émergé en tant que science multidisciplinaire dans le but d'améliorer nos connaissances sur les maladies des plantes (description des symptômes, des cycles de vie des pathogènes par exemple) ainsi que de développer des méthodes visant à pouvoir les éviter ou du moins les contrôler (Lepoivre 2003).

Un des moyens les plus fréquemment utilisés par l'homme pour combattre et contrôler ces maladies consiste en l'application de pesticides, composés toxiques ou néfastes pour les pathogènes. Cette pratique est malheureusement souvent associée à des conséquences délétères sur l'homme et l'environnement (*e.g.*, pollution des écosystèmes ou impacts sur la santé publique avec comme exemple récent le dossier « Chlordécone » aux Antilles). Il est communément admis que, les champignons phytopathogènes exhibent de part différentes caractéristiques de leur histoire de vie (régimes de reproduction mixtes, grandes tailles de populations, grandes capacités de dispersion potentielles, temps de génération courts...) un fort potentiel adaptatif (*i.e.*, une grande capacité d'adaptation face aux pressions biotiques et abiotiques auxquelles ils sont soumis) (McDonald et Linde 2002). Et en effet, l'utilisation massive de fongicides dans les agrosystèmes a fréquemment conduit à l'émergence, parfois rapide, de souches totalement résistantes. D'un point de vue évolutionniste, la résistance aux fongicides est une adaptation puisqu'elle est le résultat de la sélection, qui tend à différencier des populations qui s'adaptent à des conditions environnementales différentes (Lenormand 1998). Elle est de ce fait un exemple d'évolution contemporaine *in natura* s'exerçant à une échelle spatio-temporelle très restreinte et s'opposant ainsi à la vision courante d'une évolution lente et graduelle due à de faibles pressions de sélection (Fisher 1930). Il apparaît donc important, aussi bien d'un point de vue purement scientifique que d'un point de vue appliqué, de comprendre l'adaptation des champignons phytopathogènes à leur milieu afin de développer de nouvelles stratégies de protection assurant le contrôle de l'émergence et de la propagation des gènes de résistance aux fongicides.

À l'échelle de la population, le devenir d'un allèle mutant (*e.g.*, conférant la résistance) dépend de l'interaction entre les différentes forces évolutives: mutation, dérive génétique, migration et sélection. Dans certains cas, il a été théoriquement montré que l'effet de la migration

Disease	Location	Comments
Fungal diseases		
1. Cereal rusts	Worldwide	Frequent severe epidemics; huge annual losses
2. Cereal smuts	Worldwide	Continuous losses on all grains
3. Ergot of rye and wheat	Worldwide	Poisonous to humans and animals
4. Late blight of potato	Cool, humid climates	Annual epidemics, e.g., Irish famine (1845–1846)
5. Brown spot of rice	Asia	Epidemics, e.g., the great Bengal famine (1943)
6. Southern corn leaf blight	U.S.	Epidemic 1970, \$1 billion lost
7. Powdery mildew of grapes	Worldwide	European epidemics (1840s–1850s)
8. Downy mildew of grapes	U.S., Europe	European epidemic (1870s–1880s)
9. Downy mildew of tobacco	U.S., Europe	European epidemic (1950s–1960s); epidemic in North America (1979)
10. Chestnut blight	U.S.	Destroyed almost all American chestnut trees (1904–1940)
11. Dutch elm disease	U.S., Europe	Destroying all American elm trees (1930 to date)
12. Coffee rust		Destroyed all coffee in Southeast Asia (1870s–1880s); since 1970 present in South and Central America
13. Banana leaf spot or Sigatoka disease	Worldwide	Great annual losses
14. Rubber leaf blight	South America	Destroys rubber tree plantations

Tableau 1. Exemples de différents pathogènes de plantes et des pertes engendrées.
(Adapté de Agrios 2005)

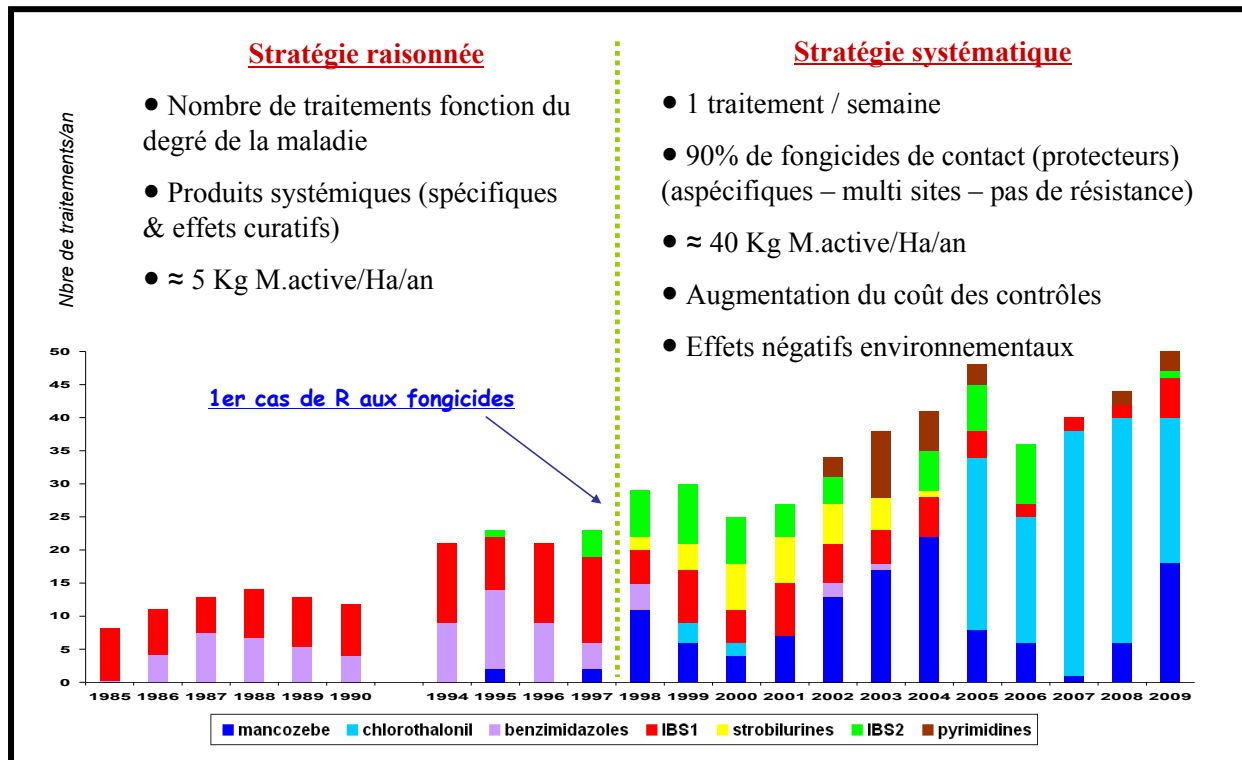


Figure 1. Evolution du nombre de traitements fongicides/an (exemple d'une plantation industrielle de la zone de production du Mounjo au Cameroun) en fonction de différents types de fongicides.

entre des zones traitées et non traitées pouvait contrebalancer l'effet de la sélection et empêcher un allèle mutant d'envahir la population traitée (Nagylaki 1975, Lenormand 2002). Ce résultat a d'ailleurs été utilisé pour le développement d'une stratégie de gestion des résistances aux insecticides chez les moustiques (Lenormand & Raymond 1998). Cependant, cette approche a rarement été appliquée au cas des fongicides. Une des raisons potentielles est que le développement d'une telle stratégie nécessite de connaître la valeur réelle des paramètres reliés à la sélection et à la migration ce qui est rarement le cas pour les champignons phytopathogènes. D'un point de vue plus général, une récente étude bibliographique regroupant la majorité des articles publiés dans le contexte théorique de l'évolution des résistances aux pesticides a montré que malgré leur importance, les flux de gènes (ainsi que la mutation et la dérive génétique) ont été trop peu souvent pris en considération dans les modèles (Rex consortium 2010).

C'est précisément dans ce contexte de résistance aux fongicides que s'insère le projet global Européen à l'origine de cette thèse. Ce projet est issu d'une collaboration entre différents organismes incluant deux centres de recherche : un Français (le CIRAD¹) et un Africain (le CARBAP²), une entreprise spécialisée dans les produits phytosanitaires (Bayer Crop Science) et l'association des producteurs de bananes au Cameroun (ASSOBACAM³). Ce projet affiche un objectif appliqué, celui d'étudier les conditions de réintroduction de certains fongicides systémiques à fort effet curatif dans les programmes de lutte contre *Mycosphaerella fijiensis*, un champignon pathogène des bananiers au sein des différentes zones de production du Sud-ouest Cameroun. Ce champignon est l'agent causal de la maladie des raies noires (MRN) présentant des conséquences économiques et écologiques néfastes majeures pour ce pays. En effet, après l'arrivée de ce parasite au Cameroun, une stratégie de lutte basée sur l'emploi raisonné de fongicides systémiques a permis de contrôler efficacement la MRN grâce à un faible nombre de traitements. Cependant, l'efficacité de ces fongicides curatifs a été annihilée par l'envahissement de populations présentant de forts niveaux de résistance. En conséquence, la stratégie de contrôle a évolué vers l'utilisation massive d'une autre catégorie de fongicides, certes ne provoquant pas l'apparition de souches résistantes mais ne présentant pas non plus d'effet curatif. Ce changement de stratégie a entraîné une augmentation du nombre de traitements, à l'origine de celle du coût économique de la lutte et des risques environnementaux associés (Figure 1). La remise en place d'une stratégie de contrôle basée sur des fongicides systémiques est donc urgente mais doit se faire en prenant explicitement en compte les facteurs impliqués dans l'évolution des résistances. Deux études de thèse ont vu le jour dans ce contexte. La première (en cours de réalisation) se focalise sur la mesure de l'intensité de la sélection exercée par différents fongicides systémiques. L'objectif de la seconde thèse (celle ci) est l'étude des flux géniques chez ce champignon phytopathogène à l'échelle du bassin de production Camerounais. À terme, les valeurs de paramètres seront intégrées dans un modèle théorique d'évolution des résistances afin de déterminer si des stratégies efficaces et durables basées sur la réutilisation des fongicides systémiques au sein des plantations Camerounaises existent.

¹ Centre International de Recherche Agronomique pour le Développement

² Centre Africain de Recherches sur Bananiers et Plantains

³ Association Bananière du Cameroun

Introduction Générale

I – Dispersion et flux géniques

I-1 – Quelques définitions

Un des phénomènes les plus étudiés (ce qui n'en fait pas l'un des mieux compris) en écologie ainsi qu'en biologie évolutive est le mouvement des individus, propagules et gènes (Clobert *et al.* 2001). Au sein du monde vivant, il existe une grande diversité de raisons et de façons d'entrer en mouvement. Par exemple, la plupart des animaux se déplacent pour se nourrir ou s'accoupler ; les végétaux et les champignons dispersent leurs graines ou leurs spores dans l'air pour survivre à des conditions défavorables ou coloniser de nouveaux habitats ; certaines bactéries et virus peuvent utiliser les voies naturelles offertes par leur hôte animal pour se transmettre à un hôte sain. Pour un scientifique qui cherche à observer et caractériser ce phénomène, il existe également différents niveaux de perception du mouvement en fonction de l'unité considérée (espèce, population, individu, gène) ainsi que différents domaines d'étude et d'application (écologie, biologie de la conservation, génétique des populations, physique). Ces différences pouvant potentiellement mener à différentes définitions et ainsi prêter à confusion dans certains cas, il est important de commencer par définir les termes qui seront utilisés tout au long de ce document. Nous appellerons « dispersion », tout mécanisme par lequel un individu se déplace de son endroit de naissance vers sa localité de reproduction (même s'il ne parvient pas à s'y reproduire). Au sens strict du terme, « migration » désigne généralement un échange d'individus entre populations. Cependant, nous l'utiliserons ici au sens large, c'est-à-dire impliquant le mouvement des gènes d'une génération à l'autre. Migration et flux de gènes seront donc ici synonymes et font référence à une dispersion dite efficace, notion très utile dans certaines disciplines de la biologie évolutive puisqu'elle permet de s'affranchir des événements qui ne participent en rien à la génération suivante. Finalement, le terme « population », concept central dans les sciences biologiques, sera fréquemment employé tout au long de ce document. De nombreuses définitions existent dans la littérature selon le paradigme auquel elles font références (Waples & Gaggiotti 2006). Ici le paradigme sera évolutif et nous appellerons population un groupe d'individus de la même espèce vivant dans une zone géographique suffisamment restreinte de telle sorte que chaque membre ait potentiellement une chance équivalente de se reproduire avec chaque autre membre du sexe opposé (Hartl et Clark 2007).

I-2 - La dispersion au cœur de la génétique des populations

La dispersion est un trait d'histoire de vie jouant un rôle central dans la dynamique et l'évolution des populations (Dieckman *et al.* 1999, Ronce 2007). Ici, je propose de regrouper les différentes implications de la dispersion en deux catégories. La première fait référence au processus d'invasion biologique c'est-à-dire à la colonisation d'un nouvel habitat alors que la seconde fait référence aux flux de gènes s'exerçant au sein et entre des populations installées. Dans les deux cas, la dispersion va avoir des conséquences sur la structure génétique spatiale des populations (SGS) (Encadré 1).

Encadré 1 : Structure génétique des populations

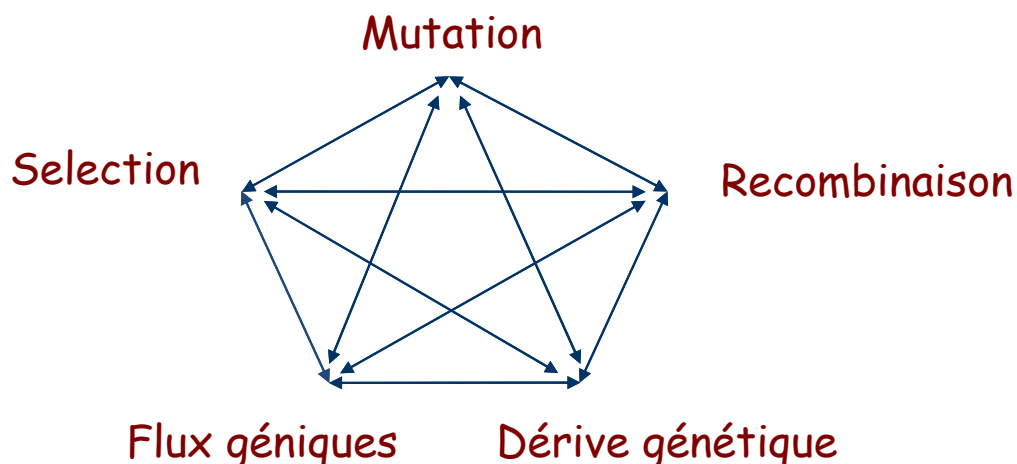
- La génétique des populations est née dans les années 1920 de la volonté de concilier la théorie Darwienne de l'évolution et les données de plus en plus précises acquises depuis le début du XX^{ème} siècle sur la transmission du matériel héréditaire. Historiquement, cette harmonisation a d'abord été faite par des mathématiciens (R.A. Fisher) et des biologistes (J.B.S. Haldane, S. Wright) qui ont construit des modèles décrivant l'évolution des populations. Un des objectifs de la génétique des populations est de décrire la structure génétique des populations.

- Les populations naturelles ont longtemps été considérées comme un ensemble homogène de gènes. Il était alors supposé que la répartition des gènes au sein de ces populations (mesurée par les fréquences alléliques et génotypiques) demeurerait d'une part la même au sein de toute la population et d'autre part constante au fil des générations. De plus, il était considéré que chacun des individus pouvait se reproduire de façon équiprobable avec n'importe quel autre individu de cette même population. Une telle population est dite panmictique. Cependant, les populations naturelles ne respectent pas forcément ces hypothèses. Elles sont souvent répandues sur de grandes surfaces, avec des barrières potentielles au mouvement des individus et avec des densités et des conditions environnementales spatialement hétérogènes. **Une population spatialement structurée présente donc des variations spatiales de répartition des fréquences alléliques.**

- Ainsi, la génétique des populations implique de raisonner en terme de fréquences des allèles dans les populations dans le but :

→ de mesurer la variation génétique dans les populations naturelles ainsi que de décrire le patron d'organisation (souvent spatial) de cette variation ;

→ d'expliquer l'origine, le maintien et l'évolution de la variation génétique par l'effet conjoint des quatre forces micro-évolutives (migration, mutation, dérive et sélection) agissant sur les individus et du système de reproduction (recombinaison).



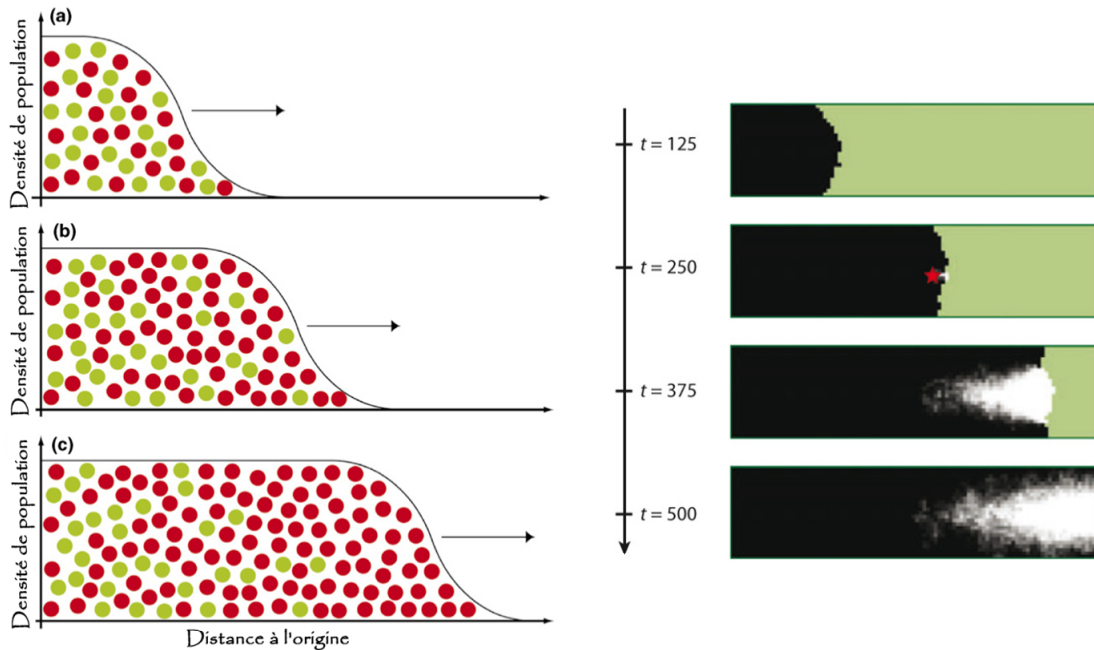
I-2-a : Dispersion et colonisation

Dispersion et colonisation sont intimement liées, puisque l'acte de coloniser un nouvel habitat ne peut se faire qu'au moyen d'un mouvement dans l'espace d'une partie du matériel génétique, autrement dit d'un événement de dispersion. La dispersion peut avoir de grandes conséquences sur la colonisation, aussi bien sur la vitesse à laquelle un matériel génétique se déploie dans l'espace (*i.e.*, la vitesse de colonisation) que sur la structure génétique spatiale mise en place (Excoffier *et al.* 2009). Il a notamment été montré que cette structure génétique se déterminait sur un petit nombre de générations (qui constituent la phase d'expansion-colonisation) et mettait un grand nombre de générations à s'estomper une fois que l'espace est rempli et que la population fonctionne à l'équilibre démographique (Austerlitz & Garnier-Gere 2003, Bialozyt *et al.* 2006). Pour cette raison, une attention particulière a été portée sur la caractérisation de la SGS mise en place lors du processus d'expansion. Par exemple, en considérant un simple modèle de dispersion discret de type « stepping-stone » (modèle dans lequel la migration est modélisée par colonisations successives de proche en proche des différents dèmes à partir d'une source unique), Austerlitz *et al.* (1997) ont montré que le processus de colonisation tout au long de l'axe d'expansion s'accompagnait d'une chute progressive de la diversité génétique. Ce patron s'expliquant par l'accumulation d'une série d'événement de dérive génétique appelés « goulot d'étranglements » ou encore « effets de fondation ». Plus récemment, il a été montré qu'au-delà d'une chute graduelle de la diversité génétique accompagnant la colonisation, certains allèles initialement rares en front d'expansion pouvaient dans certaines conditions augmenter en fréquence (par effet de dérive génétique) et envahir la population nouvellement fondée (Edmonds *et al.* 2004). Ce phénomène appelé « surf génétique » (Encadré 2) (Klopfstein *et al.* 2006) peut avoir pour conséquence la mise en place de différences de distribution des fréquences alléliques entre la zone d'origine et la zone nouvellement colonisée. Une telle « révolution génétique », mimant l'effet de la sélection, n'avait été étudiée que de façon théorique (*i.e.*, par simulations) jusqu'à ce que Hallatschek *et al.* (2007) la mettent en évidence sur des patrons d'expansions de bactéries cultivées en boîtes de pétri, apportant ainsi, la première preuve expérimentale d'un tel phénomène. Un tel processus n'a à ce jour jamais été décrit en population naturelle.

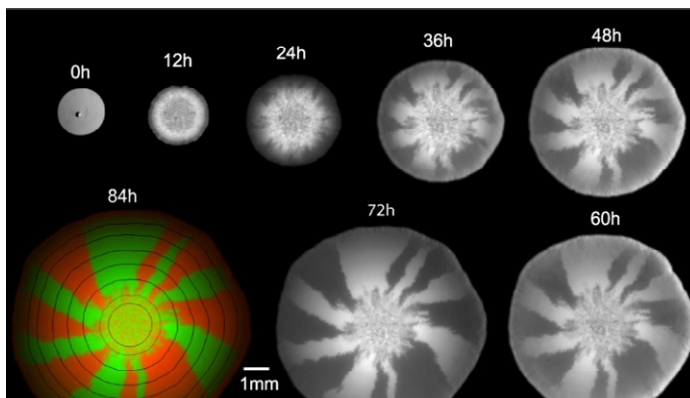
Les résultats précédemment énoncés proviennent majoritairement d'études théoriques dans lesquelles la dispersion a été simulée grâce à des modèles de réaction-diffusion. La principale limitation de ces modèles vient de la représentation de la dispersion par le terme de diffusion. Cette représentation ne permet pas une description fidèle et réaliste de la dispersion dans certains cas, notamment lorsque la dispersion à longue distance (LDD pour « long distance dispersal ») survient fréquemment. Ainsi, les modèles intégrés-différentiels ont été proposés comme alternative aux modèles de réaction-diffusion (Medlock & Kot 2003). La nouveauté, par rapport aux modèles de réaction-diffusion, réside dans un terme de dispersion qui peut prendre la forme de n'importe quel noyau de dispersion (voir I-3-a). L'utilisation de tels modèles a montré qu'une certaine quantité de migration à longue distance i) entraîne une accélération avec le temps des vitesses de colonisation (Kot *et al.* 1996), ii) diminue la probabilité d'occurrence d'un événement de « surf génétique » (Fayard *et al.* 2009) et iii) peut entraîner la mise en place d'une structure génétique spatiale en tâches (de l'anglais « patchy ») (Nichols & Hewitt 1994, Ibrahim *et al.* 1996). De telles structures,

issues de la combinaison entre événements de LDD et diffusion locale ont été récemment mises en évidence dans des populations naturelles (Parisod & Bonvin 2008, Wilmer *et al.* 2008).

Encadré 2 : Principe du phénomène de « surf génétique »



Deux Illustrations du phénomène de surf génétique. **A gauche**, schématisation du principe général de ce phénomène. Soit une population en expansion dont la fréquence de deux allèles à un locus est schématisée par des cercles rouges et verts. Dans les conditions initiales (a), les deux allèles sont équitfréquents mais l'allèle rouge se retrouve par hasard à l'extrémité de la vague et augmente en fréquence dans cette zone sous l'effet de la dérive génétique (b). Suite à cette phase de « surf génétique », l'allèle rouge a été fixé en front d'expansion (c) et le sera de ce fait également au sein de la population nouvellement fondée. **A droite**, illustration de ce phénomène de façon théorique (simulations). Au lieu de considérer qu'un allèle initialement présent dans la population envahisse la population en surfant sur la vague, les auteurs ont montré qu'une mutation neutre survenant en front d'expansion pouvait également avoir le même devenir. La colonisation de l'environnement composé de 100×25 dèmes débute du milieu de la bordure gauche de la grille et se déplace vers la droite. A un temps t arbitraire, une nouvelle mutation schématisée par l'étoile rouge survient au front d'expansion, surfe sur la vague et augmente en fréquence (zones de gris) dans le nouvel environnement. **Dans les deux cas de figure**, une rupture spatiale dans les fréquences alléliques est mise en place sous l'effet unique de la dérive génétique et de la migration. Illustrations tirée de Excoffier et al. (2009)



Première caractérisation du phénomène de surf génétique chez un organisme vivant. A $t = 0$, deux souches de bactéries (*E. coli*), différant uniquement par la nature d'un plasmide contenant des cellules marquées par deux protéines fluorescentes ont été déposées en mélange équitfréquent au centre d'une boîte de pétri. Par la suite, des photos ont été prises toutes les 12h. A $t = 84$ h, on distingue nettement la présence d'une différenciation génétique sectorielle engendrée par la ségrégation spatiale des deux allèles au locus en question due au processus de surf génétique ayant accompagné l'expansion des bactéries. Illustration tirée de Hallatschek et al. (2007)

I-2-b : Dispersion et flux géniques intra et inter populations

Une fois mis en place, les patrons spatiaux de fréquences alléliques (SGS) peuvent évoluer du fait de l'interaction entre les différentes forces évolutives (voir Encadré 1). Parmi elles, la dispersion, de par les flux de gènes qu'elle génère, a souvent été considérée comme une force limitant l'évolution des populations (Slatkin 1985). Premièrement, les flux de gènes ont pour conséquence d'homogénéiser les fréquences alléliques entre populations en les rendant plus similaires. Dans ce cas de figure, les flux de gènes s'opposent à la différenciation génétique des populations par apparitions indépendantes de mutations et/ou dérive génétique. Deuxièmement, sous effet de la sélection naturelle, les individus d'une population tendent à s'adapter à leurs conditions environnementales. Ainsi, dans un environnement spatialement hétérogène en termes de pression de sélection, on observe souvent la mise en place de patrons d'adaptation locale. Dans une telle situation, les flux géniques s'opposent à la sélection en introduisant des gènes mal adaptés aux conditions locales (créant un fardeau de migration) et peuvent limiter l'adaptation (Templeton 1986, Slatkin 1987, Lenormand 2002). En conséquence, l'effet homogénéisateur de la dispersion s'opposant à la différenciation génétique par mutation, dérive génétique ou sélection va avoir tendance à restreindre la divergence de différentes populations en différentes espèces (*i.e.* processus de spéciation) (Mayr 1963, 1970). Cependant, bien que moins souvent considérés ainsi, les flux de gènes peuvent également être vus comme une force promotrice de l'évolution des populations (*i.e.*, créatrice de diversité). En effet, les mouvements d'individus peuvent permettre la diffusion de nouveaux gènes dans des populations et ainsi participer (en interaction avec la recombinaison) à l'apparition de nouvelles combinaisons génétiques. Cela a notamment été montré de manière théorique dans le cas des métapopulations ainsi que pour des populations situées en marge des aires de distribution (Lenormand 2002).

I-3 : Mesurer la dispersion

Malgré son importance, estimer la dispersion des individus demeure un défi d'actualité en biologie du fait de la complexité de ce phénomène et des contraintes liées aux techniques utilisées pour l'étudier. En biologie des populations, on distingue généralement deux classes de méthodes permettant de mesurer la dispersion. Les méthodes « directes » ou démographiques se basent sur des observations directes du mouvement des individus. Les méthodes « indirectes » ou génétiques utilisent la distribution des fréquences alléliques pour estimer les niveaux de flux géniques ayant permis de générer le patron observé. De nombreuses revues et synthèses ont été réalisées sur les différentes méthodes existantes au sein de ces deux classes d'approches (Slatkin 1987, Bossart & Prowell 1998, Rousset 2001, 2007, Vandewoestijne & Baguette 2004, Bullock *et al.* 2006, Broquet & Petit 2009). D'autres études se sont concentrées sur la justesse, précision et puissance statistique relative des différents estimateurs de la dispersion (Slatkin & Barton 1989, Cockerham & Weir 1993, Austerlitz & Smouse 2002, Abdo *et al.* 2004, Paetkau *et al.* 2004, Berry *et al.* 2004). Ainsi, le but de cette partie, ne sera pas de faire une liste exhaustive des méthodes existantes (pour cela se référer aux références citées ci dessus) mais de présenter brièvement certaines des approches historiquement développées et/ou ayant fait l'objet d'une attention particulière au cours de cette thèse.

I-3-a : Quels paramètres de dispersion peut-on estimer ?

La dispersion est un processus aléatoire. De ce fait, la distance de dispersion est une variable aléatoire se traduisant sous la forme d'une distribution des distances parcourues. Différents paramètres peuvent être utilisés pour caractériser ce processus. Par exemple, lorsqu'on considère des populations réparties de façon discrète dans l'espace, on utilise généralement un taux de migration c'est-à-dire la proportion d'individus échangés entre deux populations. En revanche, dans le cas de population spatialement continue, la façon la plus complète de mesurer la dispersion est représentée par la distribution entière des distances de dispersion à partir d'un point donné. Une telle distribution est une fonction de densité de probabilité⁴ décrivant la probabilité qu'une propagule (en tant qu'unité de dispersion) soit déposée en un point quelconque de l'espace (Austerlitz *et al.* 2004, Kot *et al.* 1996). À partir de cette distribution, il est possible de calculer différents paramètres statistiques synthétiques (encore appelés moments de la distribution). Un moment est une façon mathématique de caractériser la distribution des distances de dispersion mais certains d'entre eux permettent de résumer (avec plus ou moins de précision) la distribution complète en quelques valeurs synthétiques (Encadré 3). Les méthodes indirectes permettent d'inférer certains de ces moments, dont le plus fréquent est σ^2 , la variance de la distribution des distances axiales de dispersion parents - descendants. L'inconvénient de σ^2 (par rapport à une distribution de distances) est qu'il ne renseigne pas sur la proportion relative d'individus dispersant à courtes et longues distances par rapports à ceux se déplaçant à des distances intermédiaires.

La dispersion à longue distance est un phénomène qui a été observé chez de nombreuses espèces végétales (Sumner *et al.* 2001) ou encore des champignons pathogènes (Brown 1997). Il existe différentes manières de spécifiquement caractériser la dispersion longue distance. La plus simple d'entre elles est la définition d'une distance seuil (Nathan *et al.* 2008) mais l'inconvénient est que cette définition du seuil est subjective et ne permet donc pas de comparer des distributions observées chez des espèces différentes. Une seconde façon de faire est basée sur l'utilisation de la kurtosis qui mesure l'aplatissement de la fonction de densité de probabilité (voir Encadré 3). La valeur de kurtosis varie entre 1 (peu de poids pour les événements longue distance) et l'infini (beaucoup d'événements longue distance). Ainsi, on parle de noyau de dispersion leptokurtique si la kurtosis est supérieure à 3 (*i.e.*, la kurtosis du noyau gaussien) c'est-à-dire que les événements à courtes et longues distances sont plus fréquents que pour un noyau gaussien, alors que les événements de dispersion à des distances moyennes le sont moins. Inversement, si le noyau de dispersion à une kurtosis inférieure à 3, on parlera de distribution platikurtique. L'avantage de la kurtosis est qu'elle n'est pas contexte-dépendante et permet donc de comparer tous les noyaux de distributions entre eux (Kot *et al.* 1996, Klein *et al.* 2006). Cependant, la kurtosis est une mesure globale de la forme du noyau et n'est donc pas seulement reliée aux événements de dispersion à longue distance. C'est pour cette raison que d'autres auteurs (Kot *et al.* 1996, Klein *et al.* 2006) ont proposé une autre caractérisation de la dispersion à longue distance en cumulant les avantages respectifs du seuil (qui est spécifique à la dispersion longue distance) et de la kurtosis. Cette caractérisation, basée sur la vitesse de décroissance du noyau de dispersion pour des distances

⁴ encore appelée « noyau » ou « fonction » de dispersion

infinies, permet de classer n'importe quelle distribution de dispersion en trois catégories : i) les noyaux à décroissance exponentielle, ii) les noyaux à décroissance plus rapide que n'importe quelle exponentielle (dits à queues légères) et iii) les noyaux à décroissance plus lente que n'importe quelle exponentielle (dits à queues lourdes).

Encadré 3 : Moment d'une distribution des distances de dispersion

- En deux dimensions, la dispersion peut être décrite par deux coordonnées (x,y) permettant de définir une distance dite vectorielle. Les coordonnées x ou y sont généralement appelés distances de dispersion axiales (e.g. notées x sur la figure) et peuvent être négatives. On peut aussi considérer une distance dite euclidienne ($\sqrt{(x^2 + y^2)}$), strictement positive par construction (e.g. notées d sur la figure).

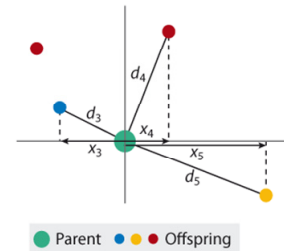


Figure tirée de Broquet et Petit 2009.

- Une distribution des distances de dispersion peut être caractérisée de façon mathématique par ses moments centrés ou non autour de la moyenne. Dans les équations suivantes, X , la distance de dispersion est une variable aléatoire.

- Moment d'ordre k non centré : $m_{k(n.c)} \equiv E[X^k] = \sum_{i=1}^n x_i^k \Pr(X = x_i)$ avec $\Pr(X = x_i) = \frac{1}{N}$

Par ex, le moment d'ordre 1 non centré : $m_{1(n.c)} = \frac{\sum_{i=1}^n x_i}{N}$ est par définition statistique une moyenn

- Moment d'ordre k centré : $m_{k(c)} \equiv E[X - E(X)]^k = \sum_{i=1}^n (x_i - \bar{x})^k \Pr(X = x_i)$ avec $\Pr(X = x_i) = \frac{1}{N}$

Par exemple, $m_{2(c)} \equiv E[X - E(X)]^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$ (égal à la variance σ^2)

- Ainsi, le moment non centré d'ordre 1 calculé à partir des distances euclidiennes correspond à la moyenne des distances parcourues.

• Cependant, en biologie on utilise préférentiellement les distances axiales pour caractériser la dispersion. Cela est très pratique car dans le cas d'une dispersion symétrique (distribution de $-x$ identique à celle de x), le moment non centré d'ordre 1 (i.e., la moyenne) est égal à 0. L'intensité de la dispersion se quantifie donc par la variance des distances de dispersion parents-descendants, équivalente à la moyenne des carrés des distances axiales parents-descendants (σ^2), cette dernière étant calculée autour du 0 moyen plutôt qu'autour d'une quelconque moyenne. Ce moment apparaît souvent dans les modèles génétiques. Puisque σ^2 s'exprime en unité de distance au carré, on utilise généralement l'écart type σ pour mesurer les distances de dispersion. En effet, dans le cas d'une distribution normale, σ correspond exactement à la moyenne des distances parcourues.

- Les moments d'ordre supérieur sont plus rarement utilisés bien qu'ils puissent être très utiles, notamment car ils permettent de mesurer une éventuelle asymétrie dans la distribution des distances de dispersion ou encore le coefficient d'aplatissement (i.e., la kurtosis)

- La kurtosis, définie comme le rapport des moments non centrés $(m_4/m_2^2)-3$ donne l'importance de la dispersion à courte et longue distance par rapport aux dispersions intermédiaires (le -3 est une convention pour que la kurtosis d'une loi normale soit nulle).

I-3-b : Les méthodes directes

Les méthodes directes (dites démographiques) sont basées sur l'observation ou la mesure de la dispersion des individus en temps réel. L'objectif de telles approches est de mesurer les distances parcourues par les individus présents sur un site donné entre leurs lieux de naissance et de reproduction. En théorie, les approches directes permettent donc d'estimer une fonction de dispersion et de caractériser les événements de LDD par ajustement des données mesurées à différents noyaux de dispersion théorique (Bullock *et al.* 2006). Cependant, dans la pratique (*i.e.*, en conditions naturelles), une telle mesure est fastidieuse pour différentes raisons, propres à la biologie et aux caractéristiques des différentes espèces. En règle générale, la plus grande difficulté est de pouvoir mesurer avec certitude la distance parcourue par un individu observé (*i.e.*, connaître son origine géographique) ainsi que de réaliser des observations à des classes de distance rarement visitées. Pour cela, les biologistes ont recours à différentes techniques comme le marquage (physique, électronique ou moléculaire) des individus. Afin d'éviter d'échantillonner tous les individus, certaines autres méthodes statistiques permettent d'estimer les mouvements globaux à partir d'un échantillon d'individus (notamment par capture-marquage-recapture, Lebreton *et al.* 2003) lorsque la biologie de l'espèce le permet. C'est d'ailleurs grâce à de telles expériences de lâchers-captures de drosophiles portant un allèle muté pour des yeux orange, que Dobzhansky et Wright (1943) ont pu étudier et mesurer pour la première fois la dispersion chez cette espèce. Lorsqu'il n'est pas possible de marquer les individus, une solution alternative consiste à rechercher leur origine *a posteriori*. Parmi les différentes méthodes développées dans ce but, les tests de parenté ont été largement utilisés (Sork *et al.* 1999, Jones et Ardren, 2003). Le succès d'une telle approche repose sur l'étape d'assignation parentale, ce qui implique un échantillonnage intensif de tous les parents potentiels ainsi qu'une puissance statistique (*i.e.*, nombreux marqueurs polymorphes) suffisante pour assigner avec certitude un génotype issu d'un individu à ses parents. Cette approche a été spécifiquement utilisée chez les plantes puisqu'elle permet (contrairement aux autres méthodes) de faire la distinction entre la dispersion de différentes propagules (*i.e.*, graines *vs.* pollen).

I-3-c : Les méthodes indirectes

L'étude de la distribution spatiale de la diversité génétique est un moyen d'obtenir des informations sur l'intensité des flux géniques en jeu. Cependant, comme nous l'avons vu précédemment, les flux géniques ne sont pas les seuls déterminants de la SGS. Ainsi, chacun des modèles théoriques présentés ci-dessous fait des hypothèses sur la nature des interactions entre les différentes forces évolutives en jeu.

Du modèle en îles en populations discrètes à l'isolement par la distance de résistance en population continue

Historiquement, le premier modèle développé dans le but d'utiliser la répartition des fréquences alléliques pour estimer un paramètre de migration a été le modèle en îles. Il a pour origine la

relation définie par Wright (1931) entre l'indice de différenciation génétique F_{ST} et le nombre de migrants dans un modèle à nombre infini de populations:

$$F_{ST} = \frac{1}{4Nm + 1} \quad \{1\}$$

où N est la taille efficace d'une population composée d'individus diploïdes et m le taux d'immigration dans cette population. L'indice F_{ST} peut être interprété comme la part de variance des fréquences alléliques qui est due aux différences entre populations, ce qui a conduit à la mise au point d'un estimateur de F_{ST} basé sur l'analyse de variance (Weir & Cockerham 1984). L'équation {1} est la relation la plus simple pour estimer des taux de migration efficaces mais cette approche a été critiquée à cause de la façon dont la dispersion est modélisée. En effet, on y fait l'hypothèse que la dispersion a lieu de façon équiprobable entre toutes les sous-populations, ce qui intuitivement semble irréaliste dans de nombreux cas (Bossart & Prowell 1998, Whitlock & McCauley 1999). Malgré cette faiblesse, le modèle en îles a été, et est toujours, largement utilisé pour comprendre les conséquences évolutives de la dispersion. Une des raisons pour cela étant que, autant les données empiriques (Bohonak 1999) que les données simulées (Slatkin & Barton 1989) indiquent que cette méthode est robuste face à des écarts potentiels aux hypothèses du modèle.

Dans le but de modéliser la dispersion de façon plus réaliste, Wright introduit dans les années 1940 le modèle dit « d'isolement par la distance » (IBD) (Wright 1943). Ce patron se traduit par une augmentation de la différenciation génétique (souvent vu comme une probabilité d'identité entre gènes) avec la distance géographique. Des résultats théoriques démontrent qu'il existe une relation linéaire entre l'indice du F_{ST} et i) les distances géographiques, pour des populations installées dans un habitat en une dimension, ou ii) leur logarithme naturel, pour des populations occupant un habitat en deux dimensions (Rousset 1997). Cette méthode a tout d'abord été développée afin de mesurer la différenciation génétique entre paires de populations discrètes :

$$\frac{F_{ST}}{1 - F_{ST}} = \frac{Q_0 - Q_r}{1 - Q_0} \quad \{2\}$$

avec Q_r la probabilité d'identité entre deux gènes mesurée entre des dèmes situés à r unités de distance l'un de l'autre et Q_0 la probabilité d'identité entre deux gènes échantillonnés dans le même dème. Elle a ensuite été étendue au cas des populations continues chez les diploïdes (Rousset 2000) par l'indice analogue du $F_{ST}/1 - F_{ST}$:

$$a_r = \frac{Q_w - Q_r}{1 - Q_w} \quad \{3\}$$

avec Q_w la probabilité d'identité entre deux gènes mesurée au sein d'un individu et Q_r la probabilité d'identité entre deux gènes échantillonnés chez deux individus séparés par la distance r . Il est important de noter que le numérateur de cette équation est estimé sur toutes les paires d'individus alors que le dénominateur est le même pour toutes les paires d'individus. Sous les hypothèses du

modèle et à l'équilibre migration-dérive, l'inférence d'un paramètre de dispersion est possible grâce à la pente de la droite de régression mesurée entre l'indice de différenciation génétique et la distance géographique :

$$b = \frac{1}{4D\pi\sigma^2} \{4\}$$

avec b la pente de la droite de régression calculée en deux dimensions (le facteur π disparaît en 1D), D la densité de la population et σ^2 la variance de la distribution des distances axiales de dispersion entre parents et descendants. Il a été montré que les estimations obtenues avec cette approche sont robustes à de nombreux écarts aux modèles sous-jacents (Leblois *et al.* 2003, 2004). Cependant, l'estimation peut être biaisée lorsque l'échelle spatiale de l'échantillonnage s'écarte sensiblement de 10σ (Rousset 2000). Ce dernier point est problématique puisque c'est justement σ que l'on cherche à estimer en utilisant ce modèle. D'autres approches différant par la statistique utilisée pour décrire la distance génétique (indices de parenté ou statistiques d'autocorrelation, Loiselle *et al.* 1995, Hardy & Vekemans 1999, Watts *et al.* 2007, Rousset 2008) ont par la suite été développées.

Finalement, le dernier raffinement du modèle d'isolement par la distance repose sur la définition d'une distance géographique autre que la distance euclidienne considérée dans les modèles précédents. En effet, la détection d'un patron d'IBD ne signifie pas nécessairement que la distance euclidienne entre populations ou individus soit seule responsable de leur différenciation génétique (Bossart et Prowell 1998). D'autres indicateurs reflétant la connectivité entre sites au sein de l'habitat peuvent apparaître plus appropriés. Dans un tel contexte, une nouvelle sorte de distance, dite de « résistance », basée sur la théorie des circuits électroniques a été proposée dans le but de prendre en compte l'intensité de la résistance d'un quelconque élément aux flux géniques (Mc Rae 2006, Mc Rae *et al.* 2007). Illustrée par la Figure 2, cette nouvelle façon de considérer la distance a permis dans certain cas d'améliorer la corrélation entre distances géographiques et distances génétiques, et de ce fait l'estimation de σ^2 . Cependant, la grosse difficulté de cette approche reste la calibration manuelle des intensités de résistance qui nécessite des données externes précises et rarement disponibles selon l'échelle considérée.

Les modèles de clines génétiques

Les modèles présentés ci-dessus ont tous pour objectif de relier un paramètre synthétique reflétant le niveau de différenciation génétique entre individus ou populations (souvent le F_{ST}) avec l'intensité de la migration. Une autre stratégie consiste à rechercher des patrons spatiaux de fréquences alléliques attendus dans certaines situations afin de faire des inférences sur les forces évolutives qui en sont à l'origine. Ce principe est à la base des modèles de clines génétiques.

Un cline génétique correspond à une variation géographique graduelle de la fréquence d'un allèle (Endler 1977). Il existe différentes raisons d'expliquer l'existence d'un patron de cline à un locus donné. La plus parcimonieuse est un contact entre populations ayant divergé suite à une phase

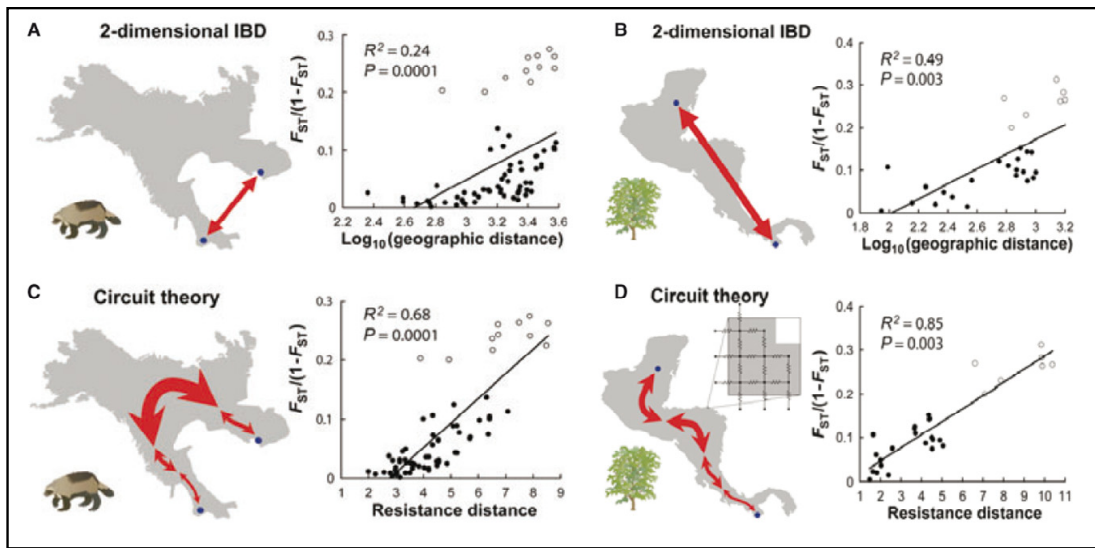


Figure 2. Illustration de l'utilisation de la distance de résistance dans un modèle d'IBD. Deux études réalisées sur des gloutons (A,C) et des arbres de mahogany (B,D) montrent que la prise en compte de la distance de résistance (C,D) améliore la corrélation entre distances génétiques et géographiques (par rapport à une distance classique euclidienne (A,B)). Les cercles blancs correspondent aux comparaisons entre paires incluant les individus du site le plus au sud du domaine. *Source : Guillot et al. (2009)*

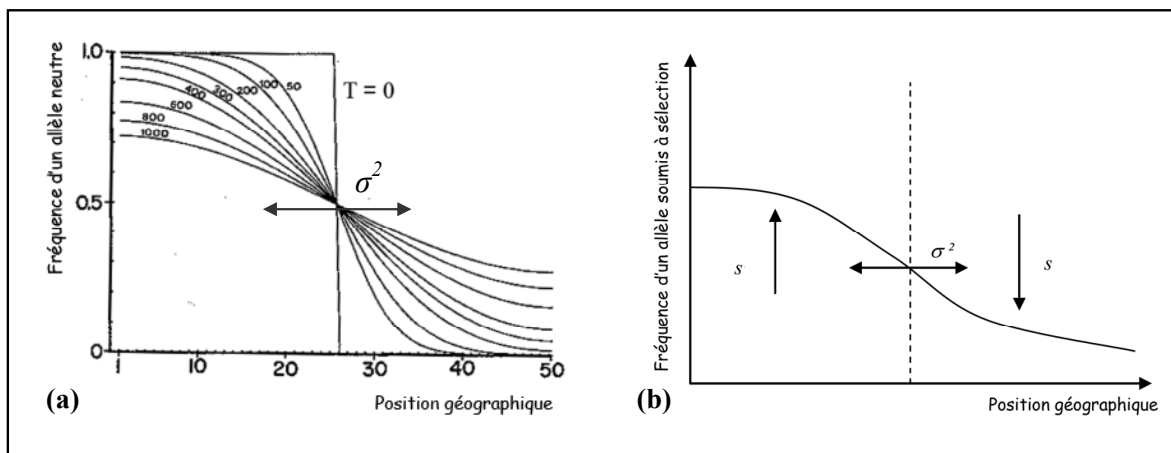


Figure 3. Représentation de deux types de patterns de clines génétiques observés : **(a)** à un marqueur neutre suite à un contact secondaire (au temps $T=0$). Un tel cline n'est pas stable, puisque la migration (σ^2) va avoir tendance à le faire disparaître (l'état du cline est ici représenté pour différents T entre 50 et 1000 générations après le contact initial). *Image adaptée de Endler (1977)*. **(b)** à un marqueur soumis à sélection. On note s l'effet de la sélection s'exerçant au sein des deux populations et σ^2 l'intensité de la migration. *Image adaptée de Labbé (2005)*.

d'isolement (*i.e.*, contact secondaire). Cette définition est la plus simple car elle n'implique que la dérive génétique (pour la fixation aléatoire de deux allèles différents au sein des deux populations revenant en contact) et la migration (qui va tendre à homogénéiser les fréquences des deux allèles par simple diffusion entre les deux populations). Un tel cline, peut donc être observé à certains marqueurs neutres (Figure 3). Cependant, il ne sera pas stable car la migration des individus va avoir pour conséquence d'homogénéiser les fréquences alléliques et donc de le faire disparaître. En théorie, en faisant l'hypothèse que les densités de populations sont identiques entre les deux populations et que les tailles efficaces sont assez grandes pour empêcher l'action de la dérive génétique, il existe une relation entre la largeur du cline (w , mesurée comme l'inverse de la pente maximale), le temps depuis le contact initial (T) et la distance de dispersion (σ , correspondant à l'écart type des distances axiales de dispersion parents-descendants ou encore la distance moyenne parcourue dans le cas d'une dispersion de type gaussienne) (Endler 1977):

$$\sigma = \frac{w}{\sqrt{2\pi T}} \quad \{5\}$$

A ma connaissance, cette situation n'a jamais été utilisée dans le but d'estimer un paramètre de migration.

L'observation d'un patron de cline génétique est également possible en considérant le même scénario de contact secondaire, et en y rajoutant l'action de la sélection naturelle. Dans ces conditions, il est possible d'observer un cline stable dans le temps suite à la mise en place d'un équilibre entre l'effet homogénéisateur de la migration et l'effet diversifiant de la sélection (voir Figure 3). Cette seconde catégorie de cline peut donc s'observer à certains marqueurs sous sélection ainsi qu'aux marqueurs neutres physiquement liés. La sélection peut être soit de type exogène (*i.e.* dépendante de l'environnement) par exemple causée par différentes pressions de sélection s'exerçant entre habitats adjacents, soit de type endogène (*i.e.* indépendante de l'environnement) comme par exemple la sélection contre les hybrides issus de la reproduction entre individus des deux populations (cette dernière étant à l'origine des « zones de tension ») (Barton & Hewitt 1985). Dans ces deux cas de figure, des modèles analytiques permettent de prédire la forme du cline (notamment l'intensité de la pente à son centre) en fonction du paramètre de dispersion σ^2 (toujours ce même moment d'ordre 2 de la distance axiale de dispersion) et d'un ou plusieurs coefficients de sélection (voir Bazykin, 1969 et Barton, 1979 pour les zones de tension et Nagylaki, 1975 pour la sélection variable dans l'espace). Par exemple, dans le cas de clines à l'équilibre soumis à sélection endogène par effet de sous-dominance (cline au sein duquel les hétérozygotes AB ont une moins bonne fitness que leurs parents homozygotes) on a :

$$\sigma = \sqrt{\frac{sw}{8}} \quad \{6\}$$

Dans ce cas de figure, la connaissance de l'intensité de la sélection (s) est donc nécessaire à l'estimation de σ^2 . Ce verrou technique a été levé au début des années 80 par les travaux de N.H

Barton sur les clines multilocus à l'équilibre dispersion/sélection. Ses développements méthodologiques ont permis d'utiliser les déséquilibres de liaison générés entre différents locus pour accéder au paramètre critique du flux génique indépendamment de la taille des populations (Barton 1983). Cette technique est basée sur le fait que la migration génère du déséquilibre de liaison (dont la valeur maximale est prédite à la zone de rencontre entre les populations divergentes) qui est ensuite maintenu par l'action de la sélection. Ainsi, la connaissance du taux de recombinaison entre les locus en déséquilibre de liaison permet d'inférer σ .

$$\sigma = w\sqrt{4Rr} \quad \{7\}$$

avec w le produit de la largeur des différents clines, R l'intensité du déséquilibre de liaison et r le taux de recombinaison entre locus.

Alors que la théorie des clines en zone hybride a été utilisée chez différentes espèces dans le but d'estimer leur dispersion (voir revue de Barton & Hewitt 1985), l'application de la théorie des clines soumis à sélection exogène est beaucoup plus récente. Deux modèles ont été développés dans le but d'étudier les clines observés à des marqueurs sous sélection (impliqués dans la résistance du moustique *Culex pipiens* aux insecticides organophosphorés) (Lenormand *et al.* 1998, 1999, 2000). Le premier s'applique à des clines à l'équilibre et nécessite au moins deux locus (pour mesurer un déséquilibre de liaison) alors que le second se base sur la variation temporelle de la pente d'un cline dans le cas de migration et/ou sélection variable dans le temps. Les deux modèles ont permis l'estimation d'un paramètre de dispersion σ par ajustement par maximum de vraisemblance des fréquences alléliques observées à des distributions attendues issues de simulations déterministes.

D'une manière générale, les modèles de clines ont été peu utilisés pour mesurer la migration (en comparaison avec les modèles « classiques » présentés précédemment), probablement parce que les situations *in natura* permettant d'observer de tels patrons ne sont pas évidentes à détecter. Cependant, ils présentent l'avantage de pouvoir analyser des dynamiques qui ne sont pas à l'équilibre, ce qui en ce sens les différencie des approches dites « classiques ».

I-3-d : Synthèse des différences entre approches directes et indirectes

Les quelques rares études ayant tenté de comparer des estimations de dispersion obtenues avec des méthodes directes ou indirectes ont permis de montrer qu'il pouvait exister des différences (Sumner *et al.* 2001, Vandewoestijne & Baguette 2004, Oddou-Muratorio & Klein 2008). Il est fort probable que ces différences soient dues au fait que les approches directes et indirectes ne mesurent souvent pas tout à fait les mêmes quantités (Bossart & Prowell 1998; Whitlock & McCauley 1999), et cela pour différentes raisons.

Tout d'abord, d'une façon générale les approches indirectes mesurent les flux de gènes tandis que les approches directes mesurent la dispersion des individus (sauf si le succès reproducteur peut être déterminé lors de la mesure directe, ce qui est rarement le cas).

Ensuite, les modèles génétiques sous-jacents aux approches indirectes font généralement l'hypothèse de neutralité des marqueurs génétiques utilisés (*i.e.*, pas de différences de valeur sélective induite par la présence des différents allèles) et d'équilibre génétique (*i.e.*, les fréquences alléliques sont constantes). Deux conséquences découlent de ces deux hypothèses: la première est que la présence de marqueurs soumis à effets sélectifs peut engendrer des erreurs dans l'estimation de la dispersion. La seconde est que l'inférence du paramètre de dispersion est couplée à un paramètre de taille efficace (reflétant l'influence relative des flux de gènes et de la dérive sur la diversité génétique). L'accès à un paramètre unique relié aux flux de gènes par les approches indirectes n'est donc généralement possible que si l'on dispose d'une mesure indépendante de la taille efficace des populations (Rousset 1997), ce qui n'est pas nécessaire avec les approches directes. Ce dernier point est problématique car la taille efficace d'une population est un paramètre très difficile à obtenir. Dans ce contexte, il est important de noter que certaines rares méthodes indirectes permettent l'estimation d'un paramètre de dispersion indépendamment de la taille efficace (les modèles de clines par exemple, voir aussi Vitalis & Couvet 2001) bien que leurs conditions d'applications restreintes limitent fortement leur utilisation.

De plus, les approches directes mesurent une dispersion localisée sur la zone d'observation et sur une période de temps souvent courte, correspondant à la durée de l'expérimentation (une ou quelques générations) alors que les approches indirectes permettent d'inférer une valeur de paramètre moyennée dans l'espace et dans le temps (pouvant correspondre à un grand nombre de générations). Autrement dit, les approches directes mesurent une dispersion actuelle alors que les approches indirectes mesurent un mélange de flux géniques contemporains et historiques. Le problème est que la part d'influence relative de ces deux types de flux géniques est dépendante du modèle génétique considéré et que très peu d'études ont été réalisées sur le sujet (à l'exception de Leblois *et al.* (2004) dans le cas de l'IBD). Les conséquences sont que les approches directes sont sensibles aux conditions liées à la période d'observation, alors que les approches indirectes sont sensibles à des variations passées de taille efficace ou de flux géniques.

Enfin, les approches directes et indirectes diffèrent par leur capacité à détecter les événements de dispersion à longue distance (Mallet *et al.* 1990, Barton & Gale 1993, Koenig *et al.* 1996, Rousset 2001). En théorie, les approches directes permettent de mettre en évidence ce genre d'événements mais le caractère « rare » les caractérisant rend leur détection beaucoup plus difficile dans la réalité. D'un autre côté, les événements de dispersion à longue distance peuvent dans certains cas influencer et biaiser l'estimation indirecte d'un paramètre unique de dispersion. Ce point a à quelques exceptions près été malheureusement très peu discuté jusqu'à aujourd'hui (mais voir Barton & Gale 1993, Lenormand 1998, Rousset 2001). Ainsi, quelques méthodes indirectes ont été développées dans le but d'inférer plusieurs moments de la distribution de dispersion (Robledo-Arnuncio *et al.* 2006, Saccheri *et al.* 2008). Ces études ont montré que même avec un effort d'échantillonnage intense, l'estimation indirecte de plusieurs moments de la distribution des distances de dispersion (dont certains caractérisant la queue de dispersion) reste un défi pour les futurs développements en génétique des populations.

En résumé, il existe de nombreuses méthodes permettant de mesurer différents estimateurs de la dispersion, mais il est important de rester prudent quand à leurs interprétations (*i.e.*, bien cerner les sources potentielles de biais dans les estimations). Nous avons vu que les approches directes et

indirectes ne mesurent pas tout à fait la même chose, ce qui les rend finalement plus complémentaires que concurrentes. Dans un tel contexte, l'analyse de potentielles différences entre estimations directes et indirectes de la dispersion devrait pouvoir permettre de mieux comprendre les processus biologiques à l'origine de telles dissimilitudes.

II – Structure génétique spatiale des populations dans la réalité des paysages

En conditions naturelles, l'inférence de paramètres reliés aux flux géniques par les méthodes indirectes présentées précédemment pose deux problèmes :

- Le premier est que la structure génétique spatiale des individus échantillonnés doit être adaptée aux hypothèses sous-jacentes du modèle théorique en question (*e.g.*, une population continue pour le modèle d'IBD, deux populations en contact pour l'approche basée sur les clines neutres, plusieurs populations discrètes pour le modèle en îles, etc). Un pré-requis à l'utilisation d'un modèle théorique est donc de connaître la structure génétique spatiale existante au sein de la zone échantillonnée.
- Le second est que, jusqu'à récemment (*i.e.*, avant l'arrivée de la notion d'isolement par la résistance), la structure génétique spatiale était modélisée comme la résultante d'une dispersion de nature isotropique⁵. Les modèles classiques ne tiennent pas compte des caractéristiques du paysage naturel dans lequel les individus dispersent. Or, dans la réalité, les paysages dans lesquels les espèces évoluent sont souvent complexes et peuvent influencer la connectivité au sein des habitats et de ce fait les flux géniques.

Dès lors, définir la structure génétique spatiale des individus (notamment dans le cas d'organismes présentant une distribution de type continue) afin de détecter des potentielles barrières aux flux géniques est de première importance avant de pouvoir utiliser un quelconque modèle inférentiel.

Dans ce contexte, de récents progrès réalisés dans le domaine des statistiques spatiales (Guillot *et al.* 2009) couplés avec le développement de méthodes issues de la génétique des populations, ont donné naissance à la génétique du paysage (Manel *et al.* 2003, Storfer *et al.* 2007, Holderegger & Wagner 2008). Cette discipline a pour but de décrire et de comprendre l'influence des structures physiques ou des caractéristiques écologiques du paysage sur la structuration spatiale de la diversité génétique. Un tel objectif se décompose en deux étapes clefs que sont i) la détection de discontinuités génétiques et ii) leur corrélation avec les variables explicatives du paysage.

Jusqu'à récemment, le moyen d'aborder ce problème consistait à prédéfinir des unités de travail sur la base d'informations obtenues *a priori* (*e.g.* origine géographique, caractéristiques phénotypiques des organismes échantillonnés) et de comparer leurs propriétés génétiques dans un cadre statistique fréquentiste. Puis, à la fin des années 90, certaines méthodes dites de classification ont été développées dans le but d'assigner des individus focaux (*i.e.* dont on ne connaît pas l'origine) à des populations prédéfinies (Rannala & Mountain 1997, Cornuet 1999). Cependant, ces approches ont

⁵ ayant la même intensité dans les différentes directions

en commun la nécessité de définir *a priori* des populations, ce qui peut être problématique lorsque les individus sont régulièrement distribués dans l'espace, et/ou lorsqu'une structuration insoupçonnée (cryptique) amène à définir des limites erronées autour de population composites. Finalement, ce verrou technique a été levé au début des années 2000 avec le développement d'une méthode Bayésienne dite de partitionnement (clustering) de type « individu centré » (Pritchard *et al.* 2000). Implémentée dans le logiciel STRUCTURE, cette approche fait l'hypothèse qu'il existe K populations (K étant inconnu) dans la zone étudiée et que chacune des populations est une unité panmictique caractérisée par ses propres fréquences alléliques aux loci analysés. La procédure d'assignation repose sur le fait que l'existence d'une structure génétique au sein d'un échantillonnage induit nécessairement un écart à l'équilibre d'Hardy-Weinberg (HW) et un déséquilibre de liaison entre loci non liés (appelé effet Wahlund). De manière simplifiée, l'algorithme utilise cet attendu théorique et permet de déterminer le nombre de populations K le plus vraisemblable au sein du jeu de données ainsi que d'assigner les différents individus aux K populations sur la base de leurs génotypes multilocus en minimisant les déviations à l'équilibre de HW au sein des populations.

Par la suite, de nouveaux algorithmes ont été implémentés dans différents logiciels et ont permis de relâcher certaines hypothèses. Par exemple, certains de ces modèles peuvent prendre en compte l'origine mélangée (*i.e.*, hybride) des individus puisque cela est souvent attendu lorsque les populations sont proches génétiquement ou ont des ancêtres communs récents dans les différentes populations (Falush *et al.* 2003). Dans ce cas, ce n'est plus l'individu lui-même qui est affecté aux différentes populations mais bien les portions analysées de son génome. D'autres manières de modéliser les fréquences alléliques entre les différentes populations (Guillot 2008) ont également permis la détection de structurations génétiques de plus en plus fines. Enfin, l'avancée la plus notable a sans doute été la prise en considération du cadre spatial de la structure des populations en introduisant une probabilité *a priori* sur les coordonnées géographiques des individus. Ce raffinement reflète le fait que les individus proches génétiquement ont généralement plus de chances d'appartenir à la même population que des individus plus distants (Guillot *et al.* 2005, François *et al.* 2006).

L'utilisation de ces différentes approches est en plein essor et suscite beaucoup d'intérêt (Guillot *et al.* 2009, Storfer *et al.* 2010, François & Durand 2010). Bien que très attrayantes, elles peuvent présenter des biais lorsque les populations s'écartent sensiblement de l'hypothèse de panmixie (Guillot 2009). Dans ce contexte, une méthode a spécialement été adaptée au cas particulier des espèces pratiquant l'autofécondation (Gao *et al.* 2007). De récents développements basés sur l'utilisation de statistiques multivariées et prenant l'espace explicitement en compte apparaissent également prometteurs puisqu'ils ne reposent sur aucun modèle génétique (Jombard *et al.* 2008, 2010). Finalement, comme toutes les méthodes indirectes, ces approches décrivent une structure génétique spatiale moyennée dans le temps. De ce fait, elles ne permettent toujours pas de mesurer l'influence relative des flux de gènes historiques et contemporains sur la SGS observée.

III- Le modèle d'étude : un champignon phytopathogène aérien dans son paysage agricole

Nous avons vu que la compréhension des flux géniques s'avère être importante d'un point de vue fondamental puisque ces derniers affectent la mise en place et l'évolution de la diversité génétique spatiale. Chez les agents pathogènes, ces processus gouvernent de ce fait respectivement la propagation des épidémies (Archie *et al.* 2009, Biek & Real 2010) et les processus d'adaptation des pathogènes (McDonald & Linde, 2002). L'étude de la dispersion et des flux géniques chez de tels organismes s'inscrit donc également dans une thématique appliquée puisqu'elle apparaît indispensable à l'amélioration des stratégies de surveillance, de quarantaine, d'éradication, de lutte contre les maladies ou encore de gestion de l'adaptation de ces micro-organismes.

III-1 : L'étude de la dispersion chez les champignons phytopathogènes

Les champignons phytopathogènes aériens dispersent de façon naturelle par dissémination de leurs organes de propagation selon trois phases pouvant se dérouler plus ou moins concomitamment (Lepoivre 2003). Tout d'abord, les spores sont libérées dans l'air selon des mécanismes biophysiques liés à la structure des organes. On distingue les mécanismes de libération actifs (lorsque l'énergie nécessaire à la libération est fournie par l'organe), ou des mécanismes passifs (lorsque l'énergie est fournie par une force extérieure). Ensuite, les spores sont transportées par voie aérienne jusqu'à la phase de dépôt (appelée pollution) à la surface de l'hôte nouvellement infecté (pouvant parfois être celui qui a libéré les spores (et on parle alors d'auto-infection)). Pour que la dispersion puisse mener aux flux géniques, il est indispensable que ces trois phases soient réalisées avec succès. Ainsi, puisque la phase de pollution a potentiellement de grandes chances de se solder par un échec, la sélection naturelle a conduit les champignons phytopathogènes à produire et libérer une très grande quantité de propagules (Brown 1997).

L'étude des processus de dispersion et de flux géniques des champignons phytopathogènes aériens peut être appréhendée à différentes échelles reflétant deux modes distincts de dispersion des individus (Brown et Hovmöller, 2002). La première forme correspond à une invasion ponctuelle par des événements rares de dissémination des parasites, soit par la dispersion naturelle et exceptionnelle de spores sur de longues distances soit par des mouvements de matériel végétal infecté. Ceci entraîne une dispersion rare et stochastique des individus plutôt aux échelles mondiale et continentale (Figure 4). La deuxième forme de dispersion est graduelle et permet l'extension d'une maladie dans un lieu donné. Elle intervient donc à une échelle plus locale et s'effectue principalement par les moyens naturels de dispersion du pathogène.

Historiquement, les premières études visant à mesurer la dispersion des champignons phytopathogènes ont été réalisées à une échelle locale grâce à des méthodes directes. Chez les pathogènes, le suivi de la progression spatio-temporelle des maladies à une telle échelle peut s'avérer très utile pour comprendre les mécanismes de dispersion sous-jacents (Jeger 1989). Cependant, chez les champignons phytopathogènes, de telles études ont rarement été réalisées sur

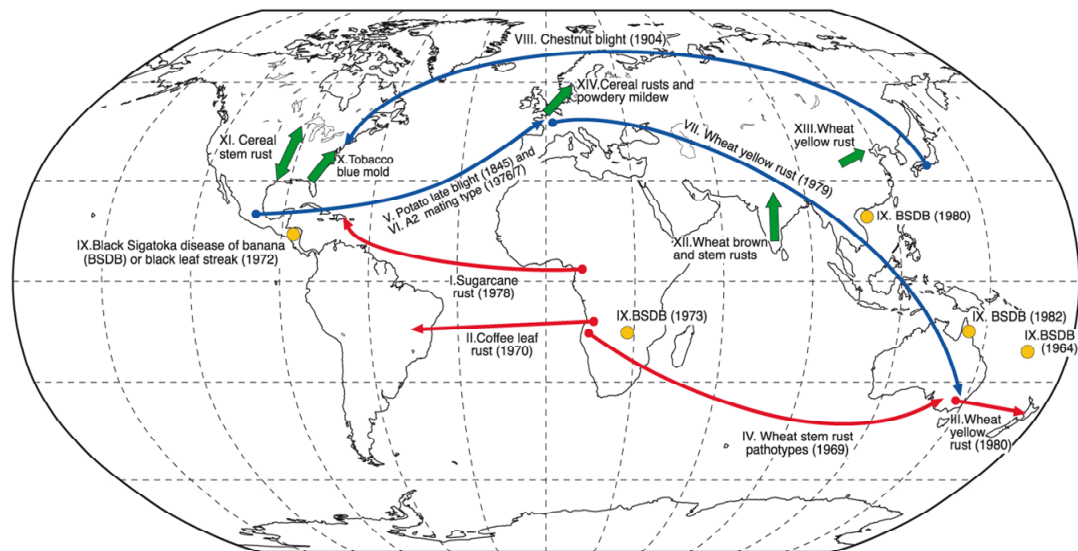


Figure 4. Exemples de différents évènements de dispersion à longue distance recensés chez des champignons phytopathogènes. Les flèches rouges et bleues indiquent des invasions de nouveaux territoires (avec l'année d'observation). Les flèches rouges indiquent des évènements de dispersion ayant probablement été réalisé par des mouvements directs de spores alors que les flèches bleues font références à ceux dans lesquels l'homme est probablement intervenu (par transport de matériel végétal infecté par exemple). Les cercles orange indiquent les dates de premières observations de *M. fijiensis* au sein des différents continents. Les flèches vertes illustrent des migrations périodiques de spores dues à des cycles d'extinction-recolonisation. Source : Brown et Hovmöller (2002).

des épidémies naturelles, probablement pour deux raisons principales. La première est qu'il faut être préparé à l'arrivée d'une maladie dans une zone bien précise pour pouvoir assurer son suivi spatio-temporel ce qui est rarement le cas, notamment car de nombreuses introductions ont eu lieu à une période pendant laquelle on ne prêtait pas ou peu d'attention aux risques associés à la maladie en question (Palm 2001). La seconde est que l'obtention de telles données est souvent techniquement fastidieuse du fait de la progression rapide des pathogènes (Agrios 2005). Ainsi, la majorité des études ayant analysé des données épidémiologiques dans le but d'inférer des paramètres reliés à la dispersion naturelle de champignons phytopathogènes ont été réalisées à partir d'inoculations artificielles et à l'échelle réduite de la parcelle (Mundt 1989, Fontem *et al.* 1991, Inman *et al.* 1999, Cox & Scherm 2001). De telles études ont permis dans certains cas de mesurer un gradient de dispersion mais l'occurrence de contaminations extérieures et/ou l'impossibilité d'échantillonner hors de la parcelle ont empêché la détection d'évènements rares de dispersion à plus grandes distances.

Par la suite, l'avènement des méthodes de reconstruction phylogéographiques⁶ en génétique des populations (Avisé *et al.* 1987) a permis la réalisation d'études aux échelles globales. De nombreuses méthodes statistiques ont été développées dans cet objectif (voir pour revue Knowles *et al.* 2009, Estoup et Guillemaud 2010) et leur application a notamment permis de retrouver l'origine ainsi que de retracer les routes d'expansions de certains champignons phytopathogènes à une échelle mondiale ou continentale (Rivas *et al.* 2004, Banke & McDonald 2005, Gobbin *et al.* 2006,

⁶ Approches dont le but est d'étudier les relations généalogiques entre populations afin d'en déduire les processus historiques et démographiques ayant mené aux structurations génétiques et géographiques observées

Brunner *et al.* 2007, Gladieux *et al.* 2008, Barres *et al.* soumis, Robert *et al.* soumis). Finalement, quelques autres études ont été menées aux échelles nationales et régionales (> 300 Km) (Milgroom & Lipari 1995, Et-touil *et al.* 1999; Breuillin *et al.* 2006, Guerin *et al.* 2007, Hayden *et al.* 2007, Montarry *et al.* 2010, Halkett *et al.* 2010). Ces études ont permis la détection de populations génétiquement structurées ainsi que de retracer les routes de colonisation à de telles échelles. Cependant, la présence de populations en déséquilibre génétique ainsi que la non-connaissance des tailles efficaces de populations n'ont pas permis l'estimation des flux géniques naturels et graduels à une telle échelle. De plus, malgré la distribution souvent continue des populations de pathogènes, la plupart de ces études ont considéré un échantillonnage en populations discrètes ce qui a empêché une localisation spatiale précise des discontinuités génétiques ainsi que leur corrélation avec des barrières potentielles aux flux géniques.

III-2 : Le pathosystème : *Mycosphaerella fijiensis* / bananier

Avec ses 110 Mt de production annuelle (Lescot 2006), la culture de la banane pointe au quatrième rang économique mondial des plantes cultivées dans un but alimentaire (après le riz, le blé et le maïs). Il existe deux grands types de bananes d'un point de vue alimentaire: les bananes douces (dites dessert) et les bananes à cuire, parmi lesquelles les bananes plantains occupent une place prépondérante (bien que très peu connues en Europe). Ces bananes sont cultivées dans de nombreux pays tropicaux selon deux schémas de production bien différents (Photo 1).

Le premier correspond à la production de bananes visant à être exportées et concerne principalement les bananes douces. Ce type de production est réalisé i) par un petit nombre de pays tropicaux (pour 80% en Amérique Latine), ii) principalement par monoculture d'un seul groupe de cultivars (quelques clones qui sont des triploïdes de l'espèce *Musa acuminata*, AAA, sous-groupe Cavendish) et iii) au sein de quelques vastes plantations industrielles. Bien qu'économiquement plus important, le mode de production voué à l'exportation ne représente que 15 % de la quantité totale produite (Arias *et al.* 2003). Le second mode de production correspond à la production de bananes d'autoconsommation. Il concerne un plus grand nombre de variétés de bananes desserts et à cuire. Il constitue une des principales ressources alimentaires de certaines populations et est réalisé au sein de nombreuses petites plantations paysannes dans plus de 120 pays tropicaux de part et d'autre du globe (Sharrock et Frison 1999).

La maladie des raies noires (MRN), causée par le champignon pathogène *Mycosphaerella fijiensis* est la maladie la plus néfaste (d'un point de vue économique) chez cette culture (Jones 2000) pour deux raisons. La première est que le développement du champignon induit une diminution de la surface photosynthétique de la plante entraînant des pertes de rendement pouvant atteindre 50 % (notamment par une baisse du poids des régimes) (Stover *et al.* 1983, Ramsey *et al.* 1990). La seconde est liée à l'aptitude du fruit à la conservation qui est une caractéristique importante des filières d'exportation. En effet, la banane est un fruit climactérique (sa maturation est dépendante de l'éthylène) ce qui explique qu'elle soit cueillie encore verte (non arrivée à maturité). L'effet principal de la MNR est une réduction importante du temps de maturation post-récolte (Stover

1972). Parce que les fruits doivent arriver encore verts dans les mûrisseries industrielles, si elle n'est pas correctement contrôlée, la MRN peut avoir des conséquences économiques très importantes sur la production de type « exportation ». Il est important de noter qu'il existe des cultivars résistants à la MRN (aussi bien pour la banane dessert que pour celle à cuire) (Jones 2000). Cependant, ces cultivars ont très rarement été utilisés pour la production de type industriel car cela obligerait l'ensemble de la filière à modifier de nombreux procédés post-récolte actuellement en place et spécifiquement adaptés à la culture des bananiers de type Cavendish (de Lapeyre de Bellaire *et al.* 2010)



Photo 1. Illustration des deux types de production de la banane : à gauche, production de type paysanne (ici banane à cuire) en bordure de route ; à droite, production vouée à l'exportation (banane dessert) au sein de grandes plantations industrielles (blocs en fond d'image) (*photographies personnelles*).

M. fijiensis est un champignon haploïde appartenant à la classe des ascomycètes dont le cycle de vie a été précédemment décrit (Stover 1972, Jones 2000) (Figure 5). Ce dernier débute par le dépôt et la germination sur une feuille d'une spore (pouvant être de type sexuée ou asexuée). Après 2 ou 3 jours de croissance épiphyllle, le tube germinatif pénètre par les stomates. En conditions optimales de température et d'humidité, les premiers symptômes apparaissent entre 10 et 14 jours après la pénétration. Ils évoluent ensuite selon 6 stades qui ont été décrits par Fouré (1982) (Photo 2). Les spores asexuées (conidies) sont produites sur les jeunes stades (stades 1 – 4) alors que les ascospores, issues de la rencontre et de la fusion entre deux mycéliums de polarités différentes sont produites dans les stades 6 (*i.e.* nécroses foliaires). Ces deux types de spores sont haploïdes puisque la fécondation est directement suivie d'une méiose. La phase diploïde est réduite à une cellule éphémère confinée à l'intérieur de la feuille.

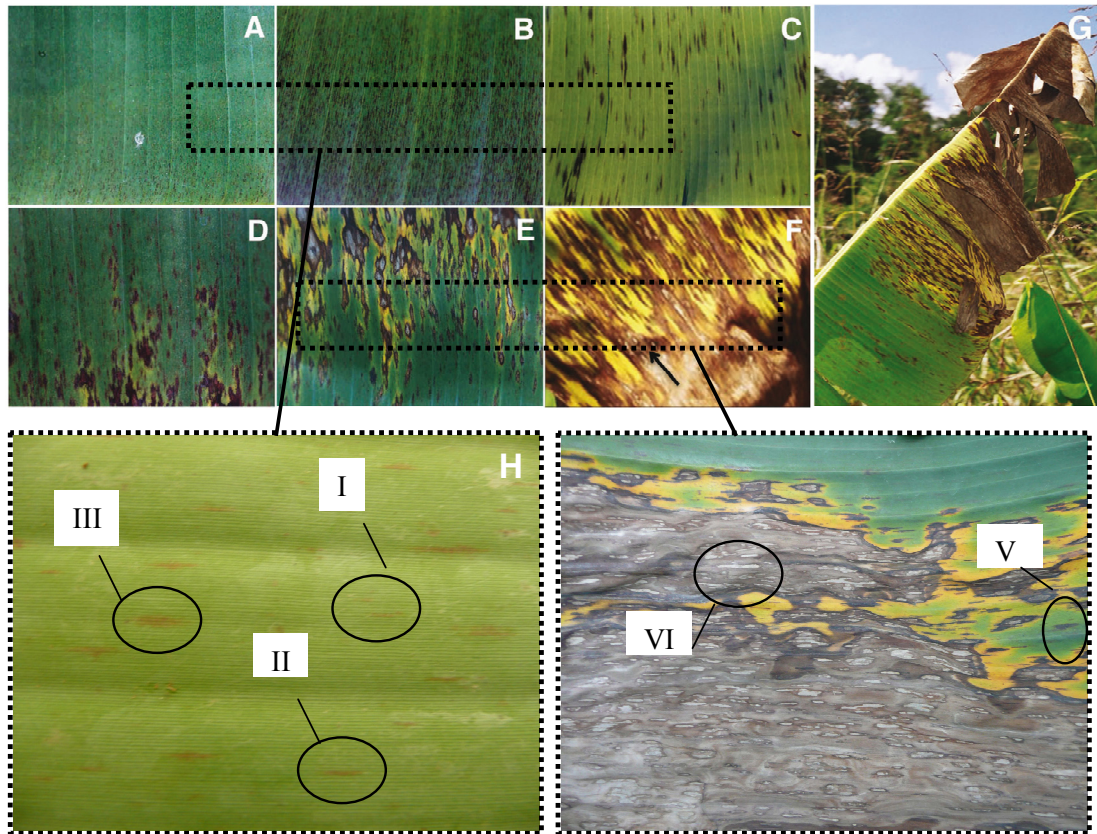


Photo 2. Différents stades de développements de la maladie des raies noires du bananier causée par *M. fijiensis*. Les stades I (H) sont difficilement visibles car les symptômes sont < 1mm. L'évolution de la maladie entraîne ensuite l'apparition de stades II (A&H) au sein duquel les lésions mesurent entre 1 et 5mm (Les lésions échantillonnées sur le terrain dans le but d'en extraire l'ADN (c.f annexe 3) sont généralement choisies à ce stade). Les lésions s'élargissent et prennent une couleur rouille pour former des stades III (B&H). La lésion atteint sa taille définitive au stade IV (C). Au stade V, les lésions noircissent (D&I) puis au stade VI, l'intérieur de la lésion devient blanchâtre (E&I). Les lésions coalescent et forment de larges plages nécrotiques au sein desquelles la production sexuée d'ascospores est réalisée (F&I). Les spores asexuées (conidies) sont, elles, produites au sein des jeunes stades (2-4). Comme illustré en (G), les différents stades peuvent être présents sur une même feuille de bananier. Photos tirées de Churchill *et al.* (2010) et de Lapeyre de Bellaire (collection personnelle).

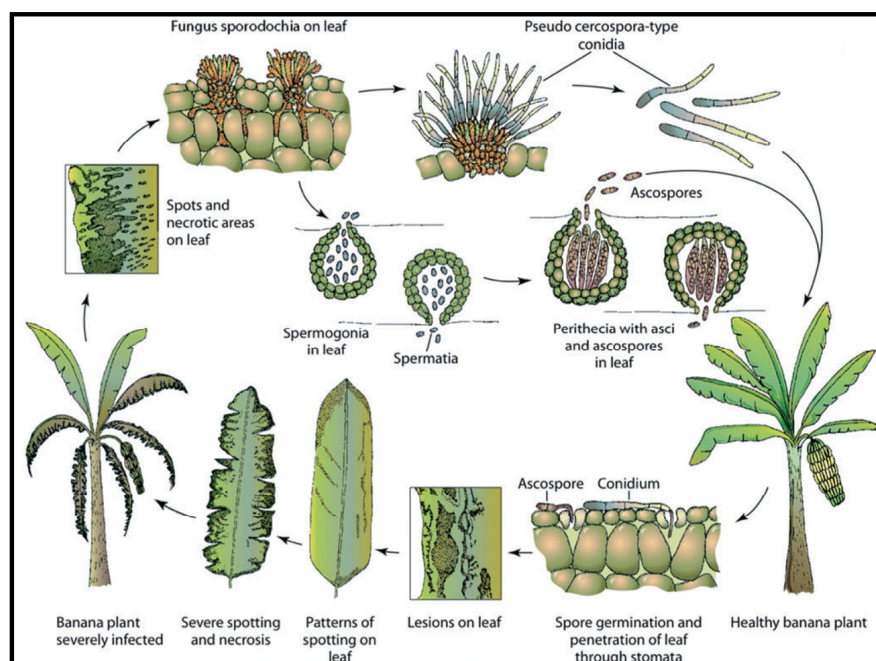


Figure 5. Cycle de vie simplifié de *M.fijiensis* sur bananier illustrant la production continue et simultanée des deux types de spores issues de la reproduction sexuée et asexuée du champignon. (Source : Churchill *et al.* 2010)

Il existe cependant une différence physique dans les modes de libération de ces deux types de spores à l'origine de prédictions divergentes sur leur potentiel de dispersion relatif. En effet, alors que les conidies présentent un mode de libération passif (elles doivent être décrochées des conidiophores qui sortent des stomates par le vent ou les eaux de pluies pour être dispersées), les ascospores sont activement libérées (*i.e.*, physiquement projetées) dans l'air après éclatement des périthèces, puis transportées par le vent. Il a d'ailleurs été montré que de telles vitesses de projection peuvent atteindre 90Km/h chez certains ascomycètes, vitesses correspondant aux accélérations parmi les plus importantes détectées dans la nature (Yafetto *et al.* 2008). Ainsi, il est généralement considéré que les ascospores participent à un transport latéral des spores pouvant potentiellement atteindre de grandes distances alors que les conidies contribuent plus à leur transport vertical à l'échelle de la plante et des plantes voisines (Mourichon *et al.* 1997). En absence de traitement fongicide, la durée du cycle de vie (entre une spore qui arrive sur une feuille et la production des spores constituant la génération suivante), bien que très dépendante des conditions météorologiques (Gauhl *et al.* 1994) est d'environ 2 semaines pour les conidies et 4 semaines pour les ascospores.

Les bananiers sont caractérisés par une émission foliaire quasi hebdomadaire qui se traduit donc par un décalage chronologique entre chaque rang de feuille. Bien que les infections primaires d'un plant puissent potentiellement avoir lieu sur toutes les feuilles, la très grande majorité d'entre elles surviennent sur les jeunes feuilles émises enroulées sur elles-mêmes (stade cigare) au sommet du bananier (Gauhl *et al.* 2000). Cela s'explique par le fait que le cigare et la plus jeune feuille déroulée (feuille 1) sont en position verticale, position physiquement plus efficace pour capter les spores présentes dans l'air que celle des feuilles suivantes. Par conséquence, il existe un gradient d'évolution de la maladie entre les jeunes feuilles du haut et celles situées dans la partie basse du bananier.

En résumé, l'émission foliaire et la culture du bananier se déroulent de façon continue toute l'année ce qui assure une réceptivité constante de l'hôte au pathogène. De plus, en zone de production paysanne, les traitements fongicides sont rares voire absents ce qui limite les fluctuations démographiques. Finalement, il a été montré que les populations sont panmictiques à l'échelle de la plante (brassage aléatoire des gamètes, voir partie suivante). Ces différentes caractéristiques font de *M. fijiensis* un bon modèle pour l'application des méthodes indirectes issues de la génétique des populations.

III-3 : Dispersion et flux de gènes chez *M. fijiensis* : Etudes antérieures

M. fijiensis est un exemple frappant d'expansion mondiale récente chez les champignons phytopathogènes. En effet, originaire d'Asie du Sud-Est (décrit pour la première fois en 1963 aux îles Fidji, (Rhodes 1964)), ce pathogène a très vite envahi toutes les principales zones de production bananières du monde (Amérique centrale, Amérique du Sud, Inde, Asie, certaines îles du Pacifique et Afrique centrale) à l'exception de l'Australie (où il a été éradiqué, Henderson *et al.* 2006) et de quelques rares îles de l'arc Caraïbéen (dont la Guadeloupe), région dans laquelle la maladie est actuellement en expansion (Abadie, com. Pers.) (voir Figure 4). Il se propage selon deux modes : le

transport de matériel végétal infecté (*e.g.* rejets infectés, feuilles malades utilisées dans l'emballage de nourriture ou autres produits), la dissémination naturelle de ces spores (Gauhl *et al.* 2000). Les deux types d'approches (directes et indirectes) ont précédemment été utilisés dans le but de décrire les différents processus de dispersion du champignon.

Des marqueurs moléculaires neutres ont été utilisés pour analyser la structure des populations de *M. fijiensis* aux échelles mondiales (Carlier *et al.* 1996, Hayden *et al.* 2003a), continentales (Rivas *et al.* 2004, Hayden *et al.* 2003b), nationales (Halkett *et al.* 2010) et à l'échelle d'une parcelle (Rivas *et al.* 2004). Les résultats montrent une diversité génétique modérée à l'échelle de la parcelle et de la plante et indiquent que les locus sont à l'équilibre gamétique dans la plupart des échantillons analysés suggérant l'existence de populations panmictiques à ces échelles. Les structures de population et les effets de fondations détectés aux échelles mondiales et continentales reflètent l'historique de la propagation de *M. fijiensis*. L'étude la plus récente (Halkett *et al.* 2010) a été réalisée à l'échelle de deux régions récemment colonisées (1979-1980) au Cameroun et au Costa Rica. Les isolats ont été collectés le long de deux transects (250/300 km) regroupant 10/15 populations échantillonnées de façon régulière (tous les 25/30 km). Des niveaux de différenciation génétique faibles par rapport à ceux observés à de plus larges échelles ont été détectés entre les différentes populations au sein des deux zones. De plus, un patron d'isolement par la distance spécifique aux populations Camerounaises a été observé. Combinés à des données historiques, ces résultats suggèrent l'occurrence d'une colonisation progressive du champignon à cette échelle régionale. Cependant, à ce jour aucun paramètre réel relié à la dispersion n'a pu être inféré.

Des approches directes ont été précédemment mises en oeuvre. Elles étaient basées sur l'analyse de gradients de maladie ainsi que sur des piégeages aériens de spores. Deux expériences de suivi du gradient de maladie à partir d'une source d'inoculum primaire ont été réalisées à l'échelle d'une petite plantation (30 x 30 mètres). La première (Abadie *et al.* en preparation), a été effectuée sur un seul cycle de vie du champignon et a permis de mettre en évidence l'existence d'un gradient de maladie abrupt suggérant une distance moyenne de dispersion par génération inférieure à 30m. La seconde (Rutter *et al.* 1998) a été exécutée durant plusieurs cycles de vie du pathogène mais n'a pas permis de mettre en évidence un gradient de maladie provenant de la source. Les conclusions de ces deux études ont en commun le fait de ne pas pouvoir exclure une « contamination » par des spores provenant de l'extérieur des dispositifs expérimentaux. Une seule étude réalisée à une plus grande échelle (entre 300m et 6000m à partir d'une source) a essayé de s'affranchir de telles contaminations extérieures en considérant une source de résistance à un fongicide (Amil *et al.* 2007). Les résultats montrent la détection de faibles niveaux de résistance à la distance maximale (6km de la source). Cependant, l'expérience a été réalisée sur plusieurs cycles de vie du champignon, ce qui ne permet pas de différencier un seul événement de dispersion à une distance donnée de plusieurs événements de dispersion cumulés à des distances inférieures. Ainsi, aucune de ces études n'a permis une estimation fiable d'un paramètre de dispersion par ce type d'approche.

Quelques expériences basées sur des piégeages aériens de spores ont tenté de clarifier l'importance relative de la dispersion des ascospores et des conidies à différentes échelles. A l'intérieur des plantations, ces études ont montré que les deux types de spores sont présentes dans l'air mais des

résultats contradictoires concernant la fréquence de capture des deux types de spores on été apportés en fonction de la région étudiée (Gauhl 1994, Burt *et al.* 1997, Burt 2003). À l'extérieur des plantations, il a été montré que les ascospores étaient capturées en fréquence plus importante que les conidies (Burt *et al.* 1998). Cette même étude a suggéré l'occurrence d'évènements de dispersion des ascospores à longues distances (entre 5 et 40 Km) mais ici encore, la probabilité de contamination par une source naturelle non détectée n'a pas pu être exclue. Finalement, une étude a montré que les ascospores de *M. fijiensis* n'étaient plus viables après 6 heures d'exposition constante à des rayons U.V. (Parnell *et al.* 1998). Bien que les spores ne soient pas continuellement exposées aux rayons U.V., cette étude semble suggérer que des transports d'ascospores viables à des distances supérieures à une centaine de kilomètres sont peu probables chez *M. fijiensis*.

En résumé, les processus naturels de dispersion des différentes propagules produites par *M. fijiensis* restent assez mal connus et surtout mal quantifiés. Il semble assez clair que les ascospores ainsi que les conidies soient dispersées par le vent. Cependant, les conidies donnent qualitativement l'impression de participer à une dispersion plus locale que celle à laquelle contribuent les ascospores même si aucune valeur de distance moyenne de dispersion n'a été mesurée jusqu'à présent. De plus, les évènements de dispersion à longue distance n'ont jamais pu être quantifiés avec certitude (car pouvant résulter d'une contamination extérieure ou de plusieurs évènements de dispersion à plus courtes distances). Enfin, les méthodes indirectes n'ont jamais été appliquées à une échelle pertinente dans le but d'inférer un paramètre relié à la dispersion naturelle ou d'étudier l'influence du paysage sur les flux géniques.

III-4- La zone de production Camerounaise du Moungo (site d'étude)

La première description de *M. fijiensis* en Afrique a été réalisée en 1973 en Zambie (Jones 2000). Au Cameroun, les données historiques relatent son arrivée en 1980 (très probablement en provenance du Sud/Gabon), date à laquelle le pathogène était confiné sur la côte Sud-Est (région de Kribi) (Fourré & Lescot 1988). En 1983, la maladie apparaît dans la zone de production bananière du Moungo et envahit progressivement tous les bananiers (plantains et desserts) de cette zone (Figure 7). Aucune information n'est disponible sur la diffusion de la maladie au Nord de cette zone de production, jusqu'à sa première description au Nigéria en 1986 (Wilson & Buddenhagen 1986). Les deux types de productions (industrielles et paysannes) sont représentés au sein de la zone. La production de type industrielle est principalement assurée par trois grosses compagnies représentant environ 6000 hectares de plantations industrielles réparties sur une dizaine de plantations. Entre ces plantations, le paysage agricole est composé de plantations paysannes de bananes plantains et desserts (petites plantations, jardins, bordures de route). Il est quasiment impossible de ne pas rencontrer ce genre de bananiers sur une surface avoisinant les quelques hectares (observation personnelle) à l'exception des zones de productions industrielles d'autres cultures (palmiers à huile, hévéa) ou de rares forêts. Une différence majeure entre ces deux types de zone de production concerne la mise en place de stratégies de contrôle de la MRN par utilisation de fongicides. Cette pratique culturale est totalement absente en production paysanne alors qu'en production industrielle, seule la lutte chimique permet actuellement l'exportation des fruits avec une durée de conservation acceptable pour la filière export (Photo 3). Dès la fin des années 80, une méthode

prévisionnelle de contrôle de la MRN (basée sur des descriptions biologiques) a été établie et appliquée avec succès limitant ainsi le nombre de traitements à 14-18 par an (Fouré, 1988) (voir Figure 1). Cette stratégie « raisonnée » reposait essentiellement sur l'emploi de fongicides systémiques à fort effet curatif. Cependant, les premières souches résistantes à certains de ces fongicides systémiques sont apparues dès 1997 et ont rapidement augmenté en fréquence (pouvant atteindre des niveaux très élevés, de Lapeyre de Bellaire *et al.* 2010). En conséquence, une nouvelle stratégie de type « systématique », reposant sur l'utilisation majoritaire de fongicides de contact (aspécifiques, préventifs) a été adoptée. Cette évolution a entraîné l'augmentation du coût économique du contrôle anti-fongique et est probablement à l'origine de nombreux effets négatifs environnementaux et sanitaires, malheureusement très peu étudiés. La raison principale est la nécessité d'appliquer ces produits lors de chaque nouvelle émission foliaire (*i.e.*, 1 fois/semaine chez le bananier) à des doses 10 fois supérieures aux doses d'utilisation des fongicides systémiques (1Kg/ha contre 100 g/ha).

En résumé, le bassin de production Camerounais du Moungo possède des caractéristiques qui lui sont propres telles i) une distribution spatialement continue du pathogène, ii) un paysage fragmenté par différents types de barrières potentielles au mouvement des individus et iii) la coexistence de plantations industrielles fortement traitées à proximité de plantations paysannes non traitées (ayant pour conséquence potentielle l'apparition de patrons d'adaptation locale ainsi que l'occurrence de flux migratoires entre ces deux zones). Pour toutes ces raisons, elle apparaît être un site intéressant pour l'étude des processus de dispersion et de flux de gènes chez *M. fijiensis*.

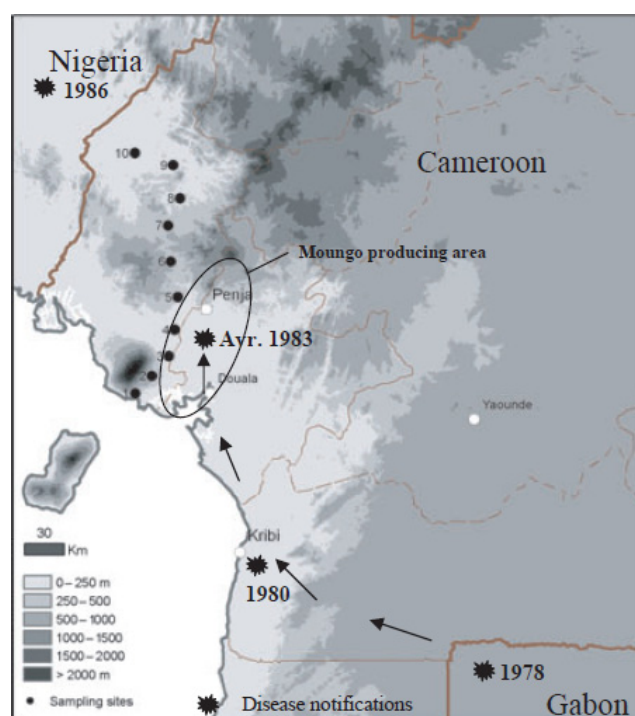


Figure 7. Colonisation historique de *M. fijiensis* au sein du Cameroun. La zone d'étude considérée pendant cette thèse est la zone de production bananière du Moungo (Sud-Ouest Cameroun). La maladie est probablement arrivée dans cette zone en provenance du Sud (Gabon). Les points noirs correspondent à l'échantillonnage réalisé par Halkett et al. (2010), article duquel à été tiré cette image.



Photo 3. Application aérienne de fongicides sur bananiers en plantation industrielle. (Lieu : Cameroun, photographe Joseph Melin)

IV- Objectifs de la thèse

Cette thèse a pour objectif l'étude des processus de dispersion et la quantification des flux de gènes qui en découlent chez *M. fijiensis* à l'échelle d'un bassin de production du Sud-ouest Cameroun. Puisque la dispersion semble être un trait d'histoire de vie difficile à mesurer de façon précise, notamment à cause de l'existence d'une importante variabilité dans les méthodes existantes, nous avons adopté une stratégie visant à combiner plusieurs approches (indirectes et directes) qui apparaissent être complémentaires dans l'échelle spatio-temporelle à laquelle elles font référence (Tableau 2). Dans un premier temps nous avons cherché à déterminer la façon dont les populations de *M. fijiensis* sont spatialement structurées à l'échelle du paysage agricole, en fonction de l'homogénéité spatiale des peuplements hôtes et de barrières potentielles aux flux de gènes (Chapitre 1). Ensuite, nous avons ciblé deux zones spatialement restreintes mais particulièrement intéressantes pour l'application des modèles de clines génétiques (Chapitre 2). Finalement, nous avons mis en place un dispositif expérimental visant à mesurer la distribution des distances de dispersion des spores chez ce champignon phytopathogène en conditions naturelles (Chapitre 3).

Suite à la présente introduction, le format de cette thèse s'articule donc autour de trois chapitres dans lesquels j'ai pris le parti d'exposer les résultats en anglais sous forme de manuscrits (qu'ils soient publiés, en cours de rédaction ou encore à un stade préliminaire d'analyses) suivis d'une discussion générale (rédigée en Français) intégrant l'ensemble des résultats obtenus ainsi que la présentation de quelques perspectives de recherche.

<div> <div> Echelle temporelle Echelle spatiale </div> </div>	Nombreuses générations	Quelques générations	Une seule génération
Paysage Agricole	<i>Structure Génétique Spatiale & Génétique du paysage (1)</i>		
Quelques kilomètres	<i>Isolement par la distance (1)</i>	<i>Clines neutres (2)</i> <i>Clines soumis à sélection (2)</i>	
Parcelle			<u>Estimation d'une fonction de dispersion (3)</u>

Tableau 2. Présentation des approches utilisées dans cette thèse dans le but d'étudier la dispersion et de quantifier les flux géniques chez *M.fijiensis* en fonction de l'échelle spatio-temporelle à laquelle elles font références. Ces approches incluent des méthodes « indirectes » (en italique) ainsi qu'une approche « directe » (soulignée.) Le chiffre en parenthèse indique le numéro de chapitre traitant de l'approche en question.

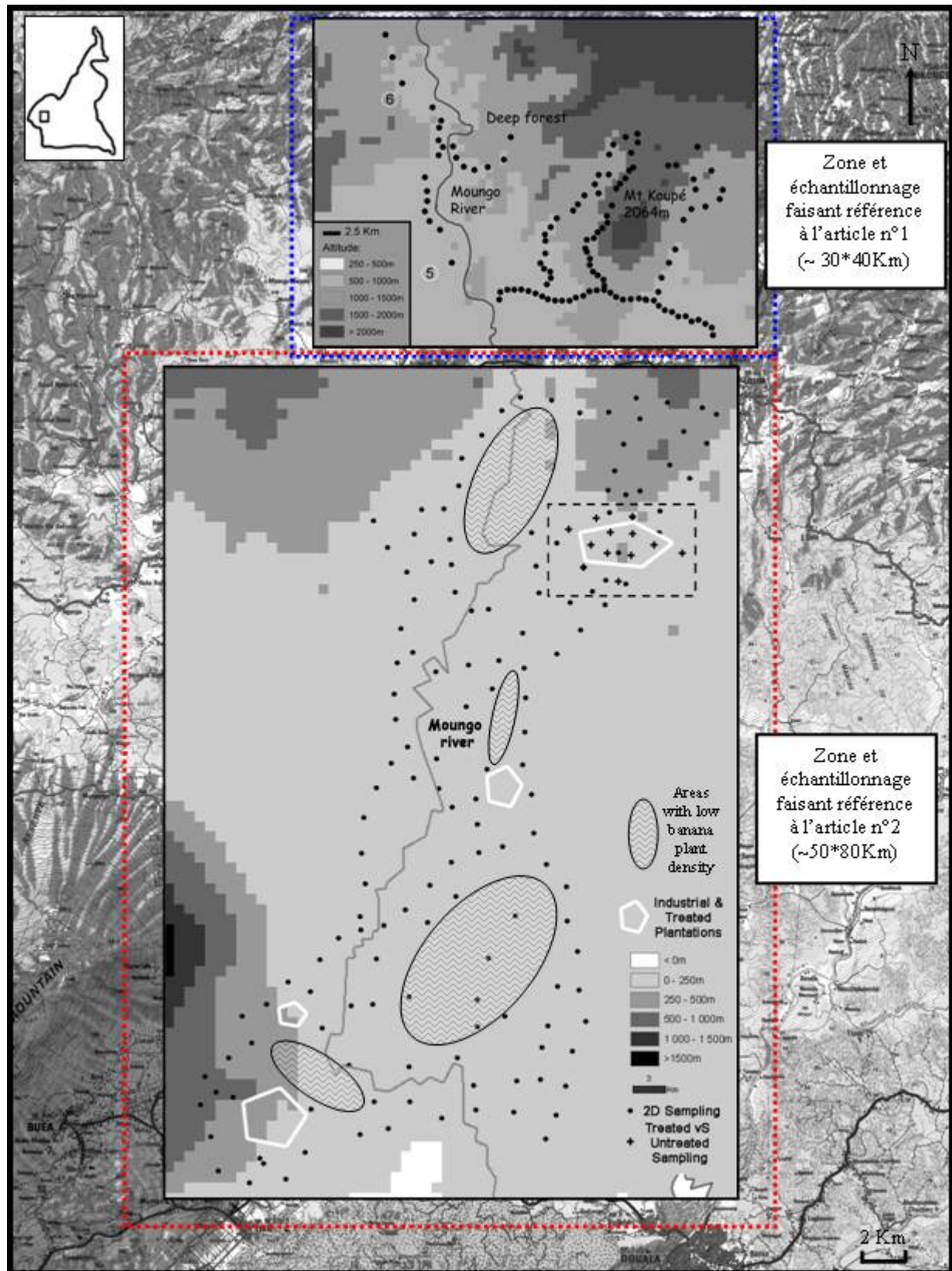


Figure 8. Paysage agricole caractérisant la zone de production bananière du Moungo (Cameroun) considérée pour les études de structuration génétique spatiale des populations.

Chapitre 1 : Structure génétique spatiale à l'échelle d'un paysage agricole

Ce chapitre repose sur deux études qui décrivent à l'aide de marqueurs génétiques neutres la structuration spatiale des populations de *M. fijiensis* à l'échelle de la zone de production bananière du Moungo. Les deux études ont été réalisées dans deux zones géographiquement proches mais présentant des barrières potentielles aux flux géniques de nature différente.

- La première étude, présentée dans l'article n°1 (Inferences on pathogenic fungus population structures from microsatellite data: new insights from spatial genetics approaches) a été réalisée au sein d'une zone d'environ 30x40km (figure 8). Dans une précédente étude incluant des échantillons de cette zone, Halkett *et al.* (2010) ont mis en évidence une différenciation génétique significative entre deux sites séparés par 30 Km (notés 5 et 6 sur la figure 8). Cependant, les deux sites échantillonnés n'avaient pas permis de spatialement localiser cette discontinuité génétique et donc de déterminer si certains facteurs du paysage pouvaient en être à l'origine. En effet, bien que cette zone se caractérise par une production bananière de type paysanne uniquement (pas d'utilisation de fongicides), le paysage agricole apparaît fragmenté puisque incluant la présence de barrières potentielles physiques aux flux géniques telles une montagne, un large fleuve ou encore une forêt dense (dans laquelle les bananiers sont rarement cultivés). Dans un tel contexte, il apparaissait important de pouvoir spatialement délimiter les populations sans avoir à faire d'hypothèses *a priori* sur leurs nombres et leurs localisations. Ainsi un échantillonnage spatialement continu en 2 dimensions a été réalisé au sein de la zone d'étude. Nous avons combiné deux méthodes d'analyses classiques de type individus-centrés prenant explicitement ou non en compte l'information spatiale et développé une procédure adaptée aux individus haploïdes permettant de décrire la structuration génétique spatiale des populations. Finalement, au sein de cette large zone, un échantillonnage a également été réalisé en 1 dimension le long d'un transect de 33km emboîté selon 3 classes de distances afin de déterminer si la distance géographique est un facteur influençant le niveau de différenciation génétique entre individus et dans le cas positif de pouvoir inférer un produit ($\sigma.N_E$). Pour l'analyse de cet échantillon, nous avons adapté le modèle d'isolement par la distance en populations continues au cas d'organismes haploïdes.

- La seconde étude, exposée au sein de l'article n°2 (No effects of recent range expansion and fungicides treatments on the spatial genetic structure of a plant pathogenic fungus in its agricultural landscape) a été réalisée au sein d'une zone située dans la continuité spatiale au Sud de celle considérée dans l'article n°1 (c.f. figure 8). Cette zone est plus vaste (50x80 km) et apparaît également plus homogène que la précédente car elle ne présente pas de barrières physiques apparentes aux flux géniques (autre que le même fleuve considéré dans l'étude précédente). Cependant, même si la majeure du paysage agricole est composée de plantations paysannes non-traitées, deux caractéristiques propres à la zone de production agricole du Moungo peuvent potentiellement réduire les flux géniques et la connectivité entre individus. Tout d'abord, quelques zones sont caractérisées par une densité très faible en hôte du fait de la production agricole d'autres cultures (hévée, palmiers à huile...). Ensuite, quatre grandes bananeraies industrielles (*i.e.*, régulièrement traitées aux fongicides) sont également présentes au sein de la zone et peuvent

potentiellement jouer un rôle de barrières chimiques aux flux géniques. L'objectif de cette seconde étude est double et fait suite au résultat obtenu dans l'article n°1 (i.e., la détection d'une discontinuité génétique d'origine historique). Dans un premier temps nous avons essayé de déterminer l'influence relative d'évènements historiques reliés au processus de colonisation du pathogène et des éléments du paysage agricole sur la diversité génétique spatiale des individus. Pour cela, nous avons réalisé un échantillonnage global en 2 dimensions sur l'ensemble de la zone en se focalisant sur des individus non traités uniquement. L'analyse de cet échantillon a été effectuée grâce à deux approches classiques d'assignation génétique. Dans un second temps, nous avons spécifiquement cherché à tester si les traitements chimiques exercés au sein des plantations industrielles pouvaient avoir un effet sur la diversité génétique neutre des populations (du fait de la dérive génétique). Pour cela, nous avons réalisé un échantillonnage hiérarchisé à l'échelle locale d'une plantation industrielle sur des bananiers traités et non traités et nous avons utilisé l'AMOVA comme méthode d'analyse.

Cet article sera prochainement soumis (revues potentiellement ciblées : Heredity, Plant pathology)

- Certaines analyses réalisées en parallèle ou à la suite de ces deux études avec les mêmes données ne figurent pas dans les articles. Cependant, certains de ces résultats me semblent importants à énoncer dans ce document car ils permettent une meilleure compréhension des conclusions rendues dans chacun des deux articles. Ils sont présentés en résultats annexes.

- L'annexe n°1 fait référence à l'utilisation d'une méthode d'inférence de paramètres démographiques (notamment de la taille efficace des populations) qui semble être adaptée au scénario historique d'expansion présumé de *M. fijiensis* dans la zone d'étude.

- L'annexe n°2 est le résultat d'une analyse d'assignation génétique spatialisée ayant été réalisée sur l'ensemble de la zone d'étude (regroupement les échantillons issus de l'article 1 + 2) grâce à certains marqueurs microsatellites en commun entre les deux jeux de données

Inferences on pathogenic fungus population structures from microsatellite data: new insights from spatial genetics approaches

A. RIEUX,^{*1} F. HALKETT,^{†1} L. DE LAPEYRE DE BELLAIRE,[‡] M.-F. ZAPATER,^{*} F. ROUSSET,[¶] V. RAVIGNE^{*} and J. CARLIER^{*}

^{*}CIRAD, UMR BGPI, Campus international de Baillarguet, TA A-54K, F-34398 Montpellier Cedex 5, France,

[†]INRA, Nancy-Université, UMR 1136 Interactions Arbre-Microorganismes, F-54280 Champenoux, France, [‡]CIRAD, UPR Syst. Banan. Ananas, TA B-26/PS4, Blvd. de la Lironde, 34398 Montpellier Cedex 5, France, [¶]Université de Montpellier II, CNRS, Institut des Sciences de l'Évolution, Montpellier Cedex, France

Abstract

Landscape genetics, which combines population genetics, landscape ecology and spatial statistics, has emerged recently as a new discipline that can be used to assess how landscape features or environmental variables can influence gene flow and spatial genetic variation. We applied this approach to the invasive plant pathogenic fungus *Mycosphaerella fijiensis*, which causes black leaf streak disease of banana. Around 880 isolates were sampled within a 50 × 50 km area located in a fragmented banana production zone in Cameroon that includes several potential physical barriers to gene flow. Two clustering algorithms and a new F_{ST} -based procedure were applied to define the number of genetic entities and their spatial domain without a priori assumptions. Two populations were clearly delineated, and the genetic discontinuity appeared sharp but asymmetric. Interestingly, no landscape features matched this genetic discontinuity, and no isolation by distance (IBD) was found within populations. Our results suggest that the genetic structure observed in this production area reflects the recent history of *M. fijiensis* expansion in Cameroon rather than resulting from contemporary gene flow. Finally, we discuss the influence of the suspected high effective population size for such an organism on (i) the absence of an IBD signal, (ii) the characterization of contemporary gene-flow events through assignment methods of analysis and (iii) the evolution of the genetic discontinuity detected in this study.

Keywords: banana, clustering analysis, emergent disease, gene flow, landscape genetics, *Mycosphaerella fijiensis*

Received 12 July 2010; revision received 24 January 2011; accepted 1 February 2011

Introduction

Understanding the forces influencing genetic diversity and quantifying the spatial scale of their influence is a central task in evolutionary biology (Dionne *et al.* 2008). Among these forces, dispersal is of particular interest for at least two reasons. First, dispersal is directly related to the colonization processes of species in new

environments. In the special case of pathogens, elucidation of the movement of individuals is crucial as such movement results in disease spread (Milgroom & Peever 2003). Second, dispersal processes determine the level of gene exchange between implanted populations, which condition their evolutionary potential (Slatkin 1987; Lenormand 2002).

Aerial fungal pathogens disperse by blowing microscopic spores into the air at high density, which renders them almost impossible to track accurately. Classically, spore dispersal is estimated by following the spatio-temporal progression of disease (Jeger 1989). However,

Correspondence: Jean Carlier, Fax: 33 (0) 499 624848; E-mail: jean.carlier@cirad.fr

¹These two authors contributed equally to the work.

for practical reasons, most of such epidemiological studies have been restricted to the parcel scale. Because agricultural landscapes are often characterized by a high degree of heterogeneity and fragmentation, it has been suggested recently that crop protection investigations and strategies should shift from the field scale to the scale of the overall agricultural landscape (Plantegegnest *et al.* 2007). Several new tools, including geographical information systems, global positioning systems (GPS), remote sensing and spatial statistics, have been developed to allow analysis and integration of the spatial components of epidemiology (Kitron 1998). Altogether, landscape epidemiology allows the essential characteristics of a landscape to be mapped and subsequently tested for any correlation between those elements and the spatial patterns of disease dynamics. Such an approach has thus far been applied mainly to insect-vector diseases (Fabre *et al.* 2005).

Landscape epidemiology may be extended by integrating tools from landscape genetics (Archie *et al.* 2009). This latter discipline combines population genetics, landscape ecology and spatial statistics to assess how landscape features and environmental variables influence gene flow and spatial genetic variation (Guillot *et al.* 2005b). A growing number of studies have been conducted in this emerging field (Coulon *et al.* 2006; Fontaine *et al.* 2007; Parisod & Bonvin 2008; Sahlsten *et al.* 2008), but until now, landscape genetic approaches have seldom been applied to pathogen populations. The new statistical tools used in these approaches are not based on populations defined a priori (Storfer *et al.* 2007) and thus appear well suited to organisms that are distributed uniformly over a given space. Landscape genetics could enhance research into landscape epidemiology because it enables the reconstruction of evolutionary relationships between parasites over a wide range of spatial scales, ranging from within individual hosts to between geographical locations (Archie *et al.* 2009). In addition to understanding the environmental factors influencing pathogen spread, this fundamentally new approach should improve our ability to track parasite movements, identify parasite origins and delineate population boundaries. In the case of pathogens, this latter point is of crucial importance because it should help identify management units.

Landscape genetics is based on the use of spatial genetics approaches that rely mostly on clustering and isolation-by-distance (IBD) models (Guillot *et al.* 2009b). Clustering approaches consider individual multilocus genotypes rather than allele frequencies to ascertain population membership or the parental origin of individuals (Rannala & Mountain 1997; Pritchard *et al.* 2000; Manel *et al.* 2005) and use clustering algorithms to sort individuals according to their genetic profile.

Hence, populations comprising individuals that regularly exchange genes can be delineated without a priori knowledge of their boundaries (Pritchard *et al.* 2000; Guillot *et al.* 2005a; Francois *et al.* 2006). A strength, but also a potential drawback, of this kind of approach is that the measure of genetic structure used to reflect gene flow is averaged over time. It is thus of prime importance to account for historical influences when interpreting contemporary genetic structures (Balkenhol *et al.* 2009). Furthermore, because dispersal is generally restricted in space for many species, continuous spatial patterns of genetic differentiation with geographical distance or IBD patterns (Rousset 2000) can induce clustering bias. Thus, the existence of IBD patterns should be tested in parallel with clustering analysis (Guillot & Santos 2009).

Mycosphaerella fijiensis, which causes black leaf streak disease (BLSD)—the most destructive leaf disease of banana (Stover & Simmonds 1987)—is a good example of a recent fungal pandemic (Mourichon & Fullerton 1990). *M. fijiensis* is a haploid and heterothallic ascomycete fungus that spreads via three modes: movement of infected plant material, and dispersal of spores produced from either sexual (ascospores) or asexual (conidia) reproduction (Gauhl 1994). Whereas conidia are dispersed mainly over short distances around the plant, wind-transported ascospores represent the predominant dispersal mode of *M. fijiensis* (Marin *et al.* 2003) and viable spores can be carried over longer distances (Parnell *et al.* 1998). Populations (groups of individuals living in close vicinity) of *M. fijiensis* are panmictic (Rivas *et al.* 2004), and their demography is relatively stable throughout the year in untreated plantation areas. Because of these latter characteristics, *M. fijiensis* is considered as a model species for fungal pathogen population genetics studies. Previous studies assessing *M. fijiensis* population structure using molecular markers have been realized at the global and continental scale (Carlier *et al.* 1996; Rivas *et al.* 2004). A more recent study examining the structure of *M. fijiensis* populations in two recently colonized areas (Halkett *et al.* 2010) showed that current population genetic structure may carry traces of the historical colonization event. In particular, a break in gene frequencies was observed along a 300-km-long transect in Cameroon. This break occurs in a hilly region. However, the coarse sampling grain in the previous sampling (performed in one dimension, every 30 km) precludes any firm conclusion that a landscape feature (e.g., a mountain) acted as a barrier to *M. fijiensis* gene flow. Alternatively, such differences in gene frequency could have arisen simply owing to demographic processes during colonization (Rees *et al.* 2009; Halkett *et al.* 2010). *Mycosphaerella fijiensis* is distributed almost continuously over space, and the discreteness of

the sampling design may have hindered the detection of a more gradual variation in allele frequencies. Thus, the use of analytical methods assuming continuous populations — together with a finer grained two-dimensional sampling scheme — should help better delineate population boundaries and assessment of *M. fijiensis* spatial population genetic structure in this area.

The main objectives of this study were to describe how *M. fijiensis* populations are structured genetically over space at the agro-system scale and to assess the influence of landscape features on spatial genetic structure. For this purpose, the genetic structure of a *M. fijiensis* population in an area of about 50×50 km in Cameroon was analysed using spatial genetics approaches. This area encompasses the genetic discontinuity previously detected (Halkett *et al.* 2010) and harbours several landscape features that could act as barriers to gene flow (e.g., mountains and forests, Fig. 1). Here, we address the following questions: (i) Is the previously detected discontinuity in gene frequencies the result of a smooth gradient owing to IBD or of a more abrupt variation? (ii) In the latter case, can this genetic discontinuity be delineated spatially? (iii) Does its location match any specific landscape feature(s) of the agro-ecosystem? In the following, particular attention has been paid to the detection, quantification and location of discontinuities in gene frequency. Genetic analyses were performed using both clustering and a novel F_{ST} -based approach. We compared clustering algorithms that either take spatial information into account or not. The F_{ST} -based approach used in this study considers continuous variation in gene frequency across the population and was adapted especially for haploid organisms.

Materials and methods

Study area and sampling strategy

In 2005, *M. fijiensis* isolates were collected from untreated plantain leaves infected with BLSD in a 50×50 km agricultural area located in southwest Cameroon. Plantains are cooking bananas [subgroup plantain, triploid (AAB) interspecific hybrid between *Musa acuminata* (AA) and *Musa balbisiana* (BB)]. The sampled area encompassed two sites (previously sampled in 2002: sites 5 and 6, Fig. 1) that displayed sharp genetic discontinuity (Halkett *et al.* 2010). Plantains are grown extensively by small farmers in the sampled area in scattered small plots without the use of any fungicide for BLSD control. Plantain (a very important food crop in this part of Cameroon) is distributed mostly around anthropic activities, i.e., residential areas, villages, roads. The crop is practically nonexistent in natural forests and at high altitudes (>1500 m). The spatial domain sampled encompasses a high mountain (Mont Koupé, 2064 m), many forests and a broad river (Mungo River). Thus, banana plantain repartition in this area is fragmented by several natural factors that could act as potential natural barriers to *M. fijiensis* gene flow. The samples were collected from easily accessible areas (roadside verges). At each sampling site, infected leaf fragments were collected from a few closely spaced banana plants, and GPS coordinates were recorded with a precision of 10 m (WGS84 data format). Between 2 and 8 *M. fijiensis* individuals were isolated per sampling site. Two specific sampling designs were used according to the type of analysis (see Fig. 1 for location and distribution of sampling sites).

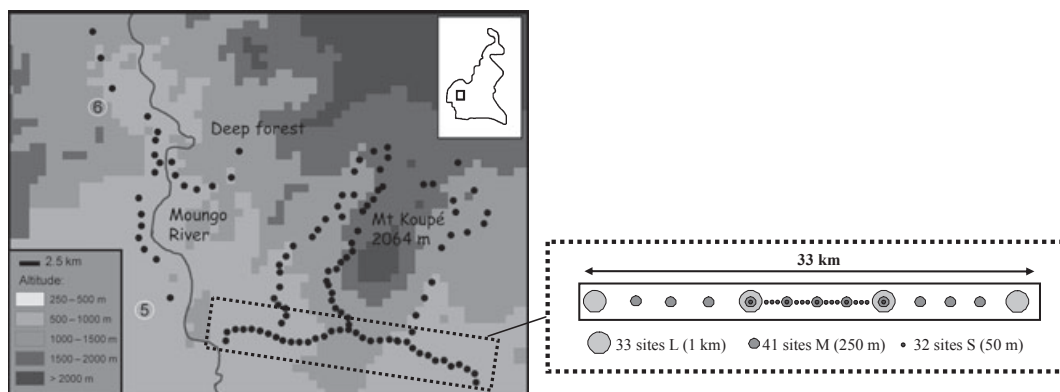


Fig. 1 Sampling distribution over the study area: *Left*: study site (50×50 km) in a region in the South-West of Cameroon. Single black dots correspond to sampled sites, white circles correspond to populations no 5 and no 6 sampled during the previous study (Halkett *et al.*, 2010). A total of 566 individuals (lesions) were sampled in 105 different sites. *Right*: Schematic representation of the 33-km-long transect showing the three classes of nested sites.

4 A. RIEUX ET AL.

Two-dimensional sampling for genetic clustering analysis. We employed a 2D sampling strategy on 105 sites spread throughout the 50 × 50 km area studied, without any a priori knowledge of population units and limits. Each site was sampled every 3 km along all roads in the spatial domain considered. A total of 567 isolates were isolated from this '2D sampling'.

One-dimensional sampling for IBD analysis. In the middle of the sampling area, 90 sites were sampled along a continuous 33-km-long transect according to three distance classes: 1 km, 250 and 50 m (embedded transect). A total of 321 isolates were isolated from this '1D sampling'. Sites more than 1 km distant (one-third of the one-dimensional sampling) were also included in the 2D sampling analysis.

Isolation procedure and molecular analysis

Mycelium cultures initiated by single ascospores isolated from necrotic lesions bearing perithecia were identified as belonging to the species *M. fijiensis* and stored as described in Carlier *et al.* (2002). Mycelium from each isolate was grown on solid medium (300 mL V8, 3g CaCO₃, 20 g agar per litre, pH 6) for 2 weeks at 25 °C, dried for 2 days at 55 °C and ground. Genomic DNA was extracted using a modification of the method described by Rogers & Bendich (1988) as detailed in Halkett *et al.* (2010). Genotyping was carried out using two polymerase chain reaction (PCR) multiplex sets comprising 17 microsatellite loci (Neu *et al.* 1999; Zapater *et al.* 2008). PCR conditions are detailed in Halkett *et al.* (2010). Finally, amplified fragments were detected and separated on a 96 capillary MegaBACE 1000 (Amersham) and analysed for length variation using GENETIC PROFILER v.1.5 Software (Amersham).

Data analysis

Basic population genetic analyses on the global sample. Population gene diversity was estimated according to the expected heterozygosity [H_E , unbiased estimate calculated according to (Nei 1978)] and the number of alleles (N_A). These two indices were calculated using FSTAT_{2.9.3} (Goudet 1995). Linkage disequilibrium between all pairs of loci was tested using Fisher's exact tests (Garnier-Gere & Dillmann 1992) implemented in GENEPOP (Rousset 2008) run on the web (<http://wbio.med.curtin.edu.au/genepop/>) [Markov chain parameters: dememorization number = 2000, number of batches = 250, number of iterations per batch = 2000]. We used the false discovery rate (FDR) procedure

implemented in the R package QVALUE (Storey 2002) to control for multiple testing. The resulting adjusted *P*-values are called *Q*-values. We measured multilocus linkage disequilibrium (\bar{r}_D) based on the index of association (Brown *et al.* 1980). This statistic is corrected for the number of loci. Departure from random association of alleles across loci was assessed by bootstrapping alleles among individuals 1000 times independently for each locus. The genotypic diversity, defined as the probability that two individuals taken at random have different multilocus genotypes, was measured. Multilocus linkage disequilibrium and genotypic diversity were calculated using MULTILOCUS (Agapow & Burt 2001).

IBD analysis. Isolation by distance was assessed using an *ad hoc* haploid version of the 'IBD between individuals' suboption available in the forthcoming version 4.1 of GENEPOP (Rousset 2008). For continuous populations, the original version considers Q_r as the probability of identity of genes between individuals at (geographical) distance r and Q_w as the probability of identity of two genes within an individual to calculate the parameter a_r so that $a_r \equiv \frac{Q_w - Q_r}{1 - Q_w}$. Here, this original model was modified for use with haploid organisms collected over a continuous spatial domain, where Q_w should be replaced by the probability of identity of genes in individuals that would compete for the same position (Robledo-Arnuncio & Rousset 2010). This cannot be estimated from real data but can be approximated by the degree of identity among the closest individuals. An analogue of the a_r parameter is then calculated to estimate the quantities $\frac{F}{1-F} = \frac{Q_r - Q_0}{1 - Q_0}$ where Q_r is the probability of identity between two genes sampled on two sites separated by a given distance r and Q_0 is the probability of identity of two genes within sites. Because the individuals in this case are haploids, at least two individuals per sampling site are necessary to estimate Q_0 .

The significance of regression slopes between genetic distances $\frac{F}{1-F}$ and the logarithm of geographical distances was evaluated in GENEPOP from Mantel tests performed across 100 000 permutations. Because the linear relationship of these variables is expected to hold best at distances greater than the mean parent—offspring dispersal distance (σ) and to progressively deviate from linearity at distances larger than 10–15 σ (Rousset 2000; Leblois *et al.* 2003), and because we do not have a real measure of σ for our pathogenic fungus, we considered a 33-km-long east—west nested transect according to three distances classes: 1 km (L = large scale represented by 33 sites), 250 m (M = medium scale represented by 41 sites) and 50 m (S = small scale represented by 32 sites; see Fig. 1). IBD was tested at

different scales on all combinations of distance classes (seven tests), on the entire data set (2D sampling) and on the two populations independently.

Assessing genetic structure using clustering methods. We applied two clustering algorithms to the 2D data set to infer population structure and assign individuals to genetic entities. First, a nonspatial algorithm, routinely used for such analysis (STRUCTURE 2.2; Pritchard *et al.* 2000), and, second, a clustering algorithm that can take spatial information into account (GENELAND 3.1.4; Guillot *et al.* 2005a). The base units of both of these methods are individual multilocus genotypes. Subdivision of the data into clusters in these methods is carried out by minimizing Hardy–Weinberg disequilibrium and gametic phase disequilibrium between loci within groups.

The STRUCTURE estimation method relies on a two-step procedure. First, the number of clusters (K) that supposedly best explains the data set is inferred from the posterior probability distribution of the data given the number of clusters, $P(X|K)$. We tested all values from $K = 1$ to 10 with 600 000 iterations of the Markov chain Monte Carlo (MCMC) algorithm following a burn-in of 60 000 iterations. We used the admixture model considering the correlated allele frequencies model but tested some runs by varying the underlying model (i.e., no admixture, and uncorrelated allele frequencies model). The number of clusters, K is calculated according to the method presented in Evanno *et al.* (2005). Moreover, for each K , we calculated the mean posterior probability over its runs and used this value to estimate the posterior probability of each K using the formula given by Pritchard & Wen (2003). Second, we ran 1 000 000 iterations with a burn-in of 100 000 iterations for the best fixed value of K . Regardless of the processing step used, five independent runs were performed to check for the consistency of Markov chain runs.

In the GENELAND procedure, all parameters (including K) are co-estimated simultaneously by the MCMC algorithm. We used both the independent and the correlated model of allele frequencies and long MCMC runs to ensure convergence of the chain. The difference between the independent (uncorrelated) and the correlated model of allele frequencies comes from the prior distributions placed on allele frequencies. The uncorrelated model turns out to give the same a priori probability to any allele frequencies (independent across loci and populations), whereas the correlated model considers that allele frequencies tend to be similar across populations (for instance, alleles that are rare in one population might also be rare in other

populations). Guillot recommended that data analysis begins with the uncorrelated frequency model, before checking how these initial results are modified by the use of the correlated frequency model (Guillot *et al.* 2009a, software manual). We followed these recommendations by performing five independent runs with 1 000 000 iterations (thinning = 100) allowing K to vary from 1 to 10 with both models. The maximum rate of the Poisson process was fixed at 500, the maximum number of nuclei in the Poisson–Voronoi tessellation at 300 and the potential error for spatial coordinates at 2 m.

The putative number of clusters was inferred from the distribution of the model value of K along the run with the highest likelihood. The run with the highest PPD (posterior probability of the data) was post-processed (with a burn-in of 5×10^4 iterations) in order to obtain posterior probabilities of population membership for each individual and each pixel of the spatial domain. We also used GENELAND without taking spatial information into account.

Spatial genetic structure assessed through an F_{ST} -based approach. We used the modified estimator of between-site genetic differentiation $\frac{F}{1-F} = \frac{Q_r - Q_0}{1 - Q_0}$ presented above to assess the spatial repartition of gene frequency differentiation between each site of the 2D sampling and a reference population (instead of considering all pairs of between-site comparisons as previously). We gathered all samples as the reference population but considered only those sites where at least three individuals have been genotyped to properly assess local gene diversity. We then mapped the variation in the local genetic differentiation estimates of $\frac{F}{1-F}$ using the interpolation MBA package implemented in the R statistical software with the aim of comparing the maps obtained with clustering analyses.

Estimates of basic statistics and genetic differentiation between clusters. The expected heterozygosity (H_E), the number of alleles (N_A), the linkage disequilibrium between all pairs of loci and the multilocus linkage disequilibrium (\bar{r}_D) were calculated within each cluster as presented in the section on 'Basic total population genetic analysis'. We used two-sided group comparisons with 1000 permutations implemented in FSTAT2.9.3 to test for significant differences in multilocus gene diversity between the two clusters. The level of genetic differentiation at microsatellite loci among clusters was estimated as F_{ST} (Weir & Cockerham 1984) using GENEPOP. We used the exact likelihood ratio (G) test implemented in this software to test for significant values of F_{ST} between inferred clusters.

Results

Whole data set genetic characterization

All 17 microsatellites were polymorphic but displayed large differences in allelic variability. The mean number of alleles ranged between 2 and 22 (5.71 on average), and gene diversity for each locus ranged from 0.02 to 0.83, with an average value of 0.40 for all loci (Table 1). Only two pairs of loci (Fb05&Fh12b and F26b&Fe05b) showed significant linkage disequilibrium after FDR control (Q -value < 0.05). Genotypic diversity was very high (0.9991), and we observed low and nonsignificant values of multilocus estimates of linkage disequilibrium ($\bar{r}_D = -0.037$). These results are consistent with the high rate of sexual reproduction of *M. fijiensis*.

Clustering results

STRUCTURE results. The estimated logarithm of likelihood for data analysed with STRUCTURE was highest for $K = 1$, and the posterior probability of having one population was equal to 1 whatever the underlying model chosen (results for admixture and the correlated allele frequencies model are presented in Table 2). For $K > 1$, none of the 567 individuals could be assigned to a given cluster as each individual had a similar probability of belonging to each cluster.

Table 1 Summary of basic information and overall population genetic analyses of *Mycosphaerella fijiensis*

Locus	N	H_E	N_A
F96	556	0.02	3
F19	567	0.03	5
F62	566	0.10	4
Ff09	548	0.10	3
N194	562	0.28	3
Fb05	534	0.28	6
F28	558	0.36	2
F26b	558	0.37	6
Ff04	558	0.41	4
N203	566	0.46	3
F40	562	0.49	3
Fe09	554	0.51	4
Fe05b	565	0.56	3
N244	555	0.64	7
N137	566	0.72	9
Ff01	520	0.72	10
Fh12b	563	0.83	22
All loci		0.40	5.71

N , sample size; H_E , unbiased estimate of gene diversity (Nei 1978); N_A , allele number, are computed over the entire sampled area among the different loci.

GENELAND results. Taking into account the spatial locations of samples in the analysis (spatial mode), we obtained slightly different results according to the underlying model of allele frequencies considered. Considering the uncorrelated model of allele frequencies (UFM), GENELAND clearly inferred the presence of two clusters (Table 2). In these runs, 543 isolates were assigned to the first cluster and 24 isolates were assigned to the second cluster (assignment threshold or posterior probability of memberships = 0.99). These two clusters appeared to be significantly genetically differentiated ($F_{ST} = 0.157$, $P < 0.001$) (Table 3) and had distinct spatial domains (Fig. 2a), with the second cluster being restricted to the extreme northwestern part of the sampling area. Considering the correlated model of allele frequencies (CFM), GENELAND inferred the presence of four clusters (Table 2). In these runs, 24 isolates (those assigned to the second cluster using the uncorrelated model of allele frequencies) were still assigned to a distinct cluster (Fig. 3a, same location as in Fig. 2a) but with a lower assignment threshold (0.90). Applying the same assignment threshold, most of the remaining individuals (450 isolates) were gathered in another cluster, located at the east side of the sampling area (Fig. 3d). These two clusters are largely differentiated ($F_{ST} = 0.204$, $P < 0.001$). The last few individuals can be confidently assigned to the two remaining clusters,

Table 2 Results of individual-based Bayesian clustering analyses computed from $K = 1$ to 10. STRUCTURE analyses: mean values of likelihood over 10 repetitions, Evanno's ΔK parameter and posterior probability of having K clusters. GENELAND analyses: mean posterior density distribution (in %) of the number of clusters estimated over the five replicates according to the different modes considered. Bold values correspond to the best K obtained in each analysis

K	STRUCTURE			GENELAND		
				Spatial		Non Spatial
	Mean $\ln P(D)$	ΔK	$\Pr(K=)$	UFM	CFM	UFM
% runs (K)						
1	-4256.4		1	0.02	0	0
2	-4343.2	30.85	2.01E-38	60.00	0	89.07
3	-4929	3.10	7.82E-293	23.33	21.06	10
4	-4802.16	4.30	9.53E-238	10.42	41.00	0.9
5	-4970.64	1.90	0	4.44	19.09	0
6	-5160.34	1.93	0	1.29	6.52	0
7	-5337.74	1.67	0	0.40	1.69	0
8	-5392.7	2.16	0	0.09	0.59	0
9	-5337.1	2.24	0	0.02	0.10	0
10	-5176.2	1.75	0	0	9.95	0.03

CFM, correlated model of allele frequencies; UFM, uncorrelated model of allele frequencies.

Table 3 Summary of basic information and genetic analyses realized on the two clusters of *Mycosphaerella fijiensis* inferred by the GENELAND analysis when considering a spatial and uncorrelated allele frequencies model

Locus	Cluster 1 (<i>N</i> = 543)		Cluster 2 (<i>N</i> = 24)		<i>F</i> _{ST}
	<i>H</i> _E	<i>N</i> _A	<i>H</i> _E	<i>N</i> _A	
F96	0.04	3	0.00	1	−0.012
F19	0.06	5	0.00	1	−0.012
Ff01	0.68	10	0.78	4	−0.007
Fh12b	0.79	21	0.87	9	0.003**
Fb05	0.30	5	0.25	4	0.014**
F62	0.19	4	0.00	1	0.034
Ff09	0.05	3	0.16	2	0.061*
N203	0.52	3	0.34	2	0.095**
Ff04	0.51	4	0.23	2	0.127**
N194	0.45	3	0.00	1	0.168***
N137	0.55	9	0.78	7	0.178***
N244	0.44	7	0.72	5	0.211***
F28	0.49	2	0.08	2	0.218***
Fe05b	0.61	3	0.30	3	0.250***
F40	0.45	3	0.29	2	0.353***
Fe09	0.38	4	0.42	2	0.369***
F26b	0.07	6	0.54	3	0.621***
All loci	0.39	5.59	0.34	3.00	0.157***

N, sample size; *H*_E, unbiased estimate of gene diversity (Nei 1978); *N*_A, allele numbers, are computed over the two clusters with related *F*_{ST} values (Fisher's exact test: **P* < 0.05;

P* < 0.01, *P* < 0.001) among the different loci.

which appeared to be somewhat intermediate between the two contrasted genetic entities (Fig. 3b, c). These two additional clusters may represent additional subtle substructuring as revealed by lower differentiation levels (all pairwise *F*_{ST} < 0.10). Ignoring the spatial locations of samples (nonspatial mode) and considering the

uncorrelated model of allele frequencies, GENELAND inferred the presence of two populations (Table 2) and the map of the posterior probabilities of population membership (after interpolation using the MBA procedure) was similar to that obtained with the spatial mode (results not shown).

*Spatial genetic structure assessed through an *F*_{ST}-based approach*

F/(1−*F*) statistic interpolation. In parallel to clustering analysis, the interpolation of local genetic differentiation into the reference population allows the spatial variation in gene frequencies to be represented (Fig. 2b). Three sites located in the uppermost left part of the map stood out, as they displayed the highest value of genetic differentiation compared to the rest of the sampling domain. Conspicuously, the localization of this discontinuity matched that detected using GENELAND.

F/(1−*F*) and 1−*Q*₀ statistics distribution around the genetic discontinuity. To better characterize the genetic discontinuity near this area, we further analysed the distribution of the *F*/(1−*F*) statistic and within-site gene diversity 1−*Q*₀ of the sites around the genetic discontinuity (Fig. 4). In this graph, each point represents a site, and the four sites at higher latitude (at the right of the graph) correspond to sites with individuals assigned to the second cluster by GENELAND using the UFM (in the northwest of the map). We can see that values of the *F*/(1−*F*) statistic are homogeneous before the discontinuity and increase directly just after the discontinuity. This result suggests that the genetic discontinuity is sharp rather than gradual. Moreover, we can see that values of 1−*Q*₀ are lower for sites located

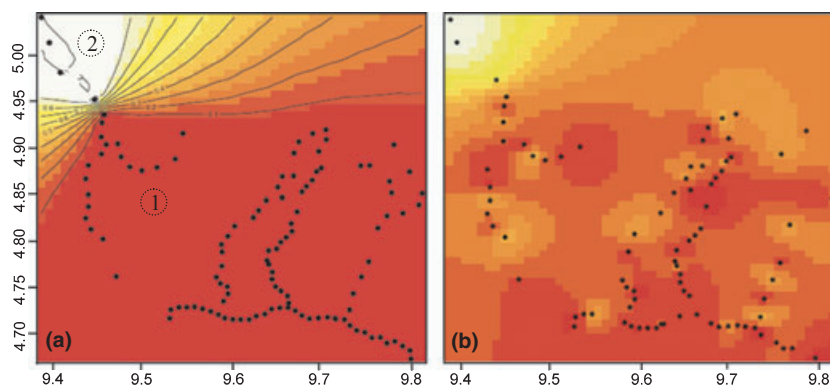


Fig. 2 Spatial genetic structure of *Mycosphaerella fijiensis* populations over the study area: (a) Map of posterior probability of belonging to cluster 1 obtained with GENELAND considering a spatial uncorrelated allele frequencies model (left); (b) Map of the genetic differentiation (*F*/(1−*F*)) parameter interpolated over the spatial domain (*R*) (right). Values are represented by a white-to-red graded scale; black dots represent sampled sites.

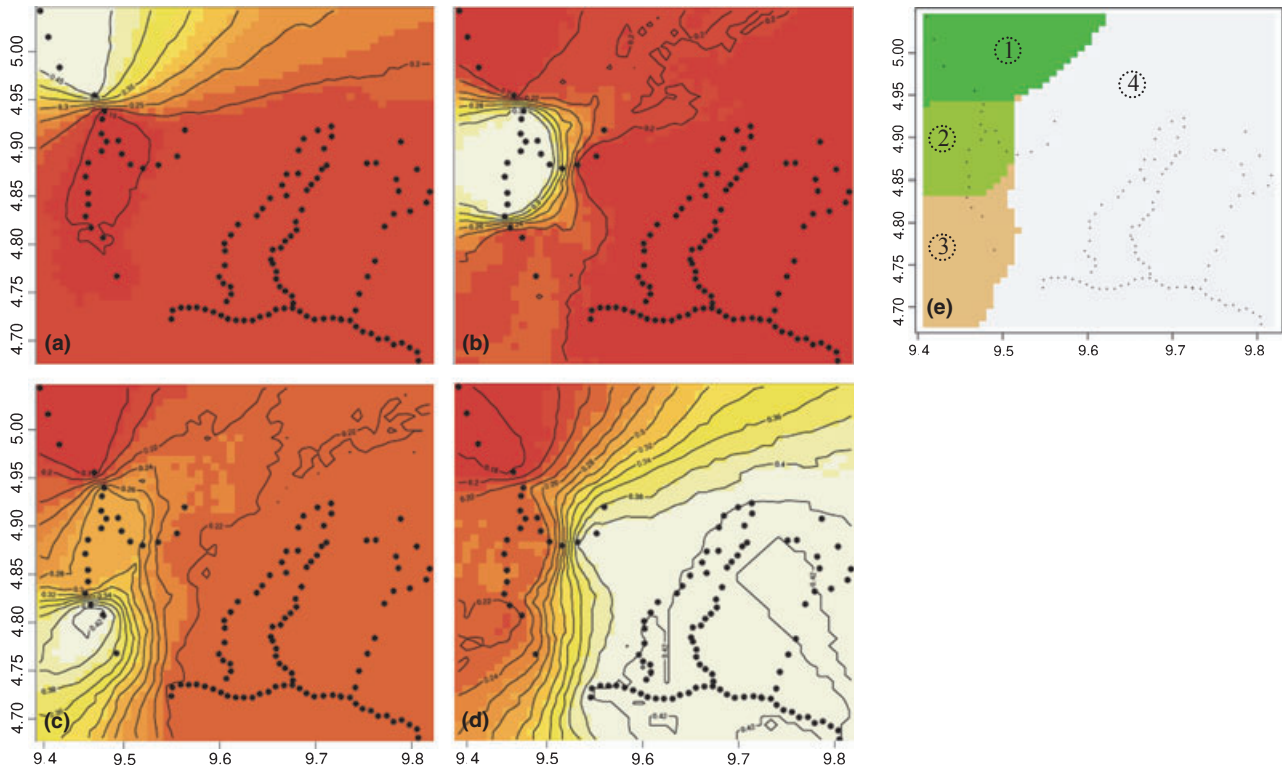


Fig. 3 Maps of posterior probability of belonging to clusters 1 (a), 2 (b), 3 (c) and 4 (d) and posterior mode of population membership (e). Results obtained with GENELAND considering a spatial and correlated allele frequencies model. Values are represented by a white-to-red graded scale; black dots represent sampled sites.

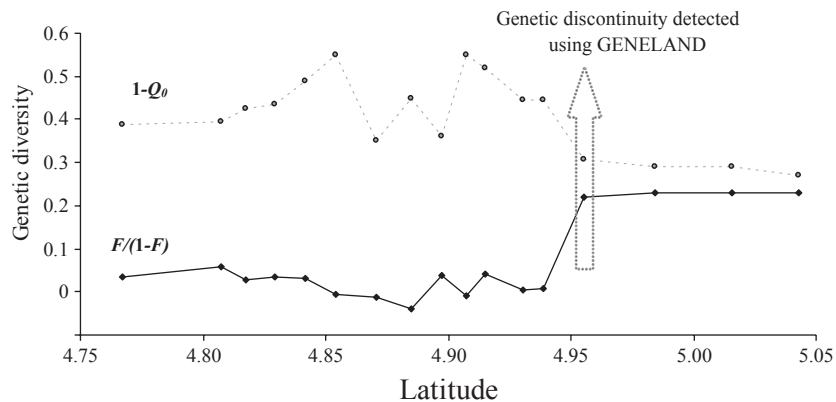


Fig. 4 Relationship between the genetic diversity measured along the site around the genetic discontinuity and latitude. The genetic differentiation inter-site parameter $F/(1-F)$ is represented by the solid line, whereas the intra-site genetic differentiation parameter ($1-Q_0$) is represented by the dashed line.

beyond the discontinuity. This result shows that allelic frequencies are homogeneous in the genetic entity located in the north. However, the high values of $1-Q_0$ detected in some sites around the discontinuity reveals a high genetic variance between individuals in these sites. This high variance that is not attributable

to smaller sample size at these sites could be explained by the presence of mixed genotypes coming from two genetic origins. Such a structure could explain the detection in this area of the two supplementary and intermediate clusters when using GENELAND with a correlated model of allele frequencies.

Table 4 Isolation by distance (IBD) within the study area according to different scales (see text for more explanation). Minimal and maximal distances, number of individuals, regression slopes with confidence interval and P -values for Mantel test of IBD are computed. Significant P values are in bold (considered level of 0.05)

Scale	Min–Max distances (Km)	n	Slope $b=$	ABC bootstrap results : Confidence interval	Test of isolation by distance (One tailed P value):
S (1 site each 50 m)	0.05–1.7	114	0.012	–0.008–0.042	0.165
M (1 site each 250 m)	0.2–9	138	0.004	–0.006–0.020	0.212
L (1 site each Km)	1–29	126	0.005	–0.007–0.026	0.344
S + M	0.04–9	238	0.003	–0.014–0.021	0.283
M + L	0.2–39	221	0.008	0.001–0.033	0.272
S + L	0.04–29	244	0.005	–0.009–0.022	0.304
S + M + L	0.04–29	321	0.007	–0.005–0.023	0.213
Entire 2D data game	0.05–59	567	0.028	0.016–0.041	0.001
Cluster 1	0.05–59	543	0.004	–0.005–0.011	0.280
Cluster 2	1–6	24	0.034	0.032–0.136	0.168

Isolation by distance. The test for IBD was significant when considering all samples, but the IBD signal disappeared when samples from cluster 1 and 2 were considered independently (Table 4). The significant IBD detected over the whole sampling domain would thus seem to stem from the second genetic cluster being located in the northwest corner of the area under study. Thus, many pairs of points that are distant at a genetic level are also distant at a geographical level.

When focussing on the 33-km-long transect situated inside the first cluster, the Mantel test of IBD was not significant whatever distance class was taken into account (Table 4).

Genetic characteristics of the inferred clusters

Given the agreement for $K = 2$ considering the spatial or nonspatial mode in GENELAND (uncorrelated model of allele frequencies) and the good concordance with the genetic differentiation analysis, we will present here the genetic characteristics of only the two main genetic entities (i.e., not considering the additional substructuring detected by GENELAND using the correlated model of allele frequency).

Multilocus values of genetic diversity (heterozygosity H_E) did not reveal any significant difference between the two clusters ($P = 0.341$). Large and significant genetic differentiation between the two clusters was observed ($F_{ST} = 0.157$, $P = 0.00002$). Again, F_{ST} values and significance varied greatly among microsatellite loci (Table 3). No pairs of loci appeared to be linked within cluster 1, but one pair (Fb05&Fh12b) showed significant linkage disequilibrium after FDR control (Q -value < 0.05) in cluster 2, maybe because of the small size of this population ($N = 24$). We observed low, nonsignificant values of multilocus estimate of linkage disequilibrium in both clusters ($\bar{r}_D = 0.041$ and 0.001 for clusters 1 and 2, respectively).

rium in both clusters ($\bar{r}_D = 0.041$ and 0.001 for clusters 1 and 2, respectively).

Discussion

We used two complementary spatial genetics approaches to study the dispersal and population genetic structure of the fungal pathogen *Mycosphaerella fijiensis* within a production area (50×50 km) in Cameroon. To our knowledge, this is the first study in which this methodology has been applied to a fungal pathogen. Our results revealed the presence of two highly differentiated populations in the study area. This genetic structure is in accordance with the results obtained in a previous study, in which Halkett *et al.* (2010) suspected a discontinuity in gene frequencies to occur along a 300-km-long transect in Cameroon. Because sampling sites were separated by approximately 30 km, it was not possible to precisely localize this discontinuity or to determine its strength. In the present study, we confirmed the presence of this genetic discontinuity and furthermore assessed the boundary of the genetic entities. It is interesting to note that the genetic discontinuity observed in 2001 can still be reported 4 years later and that the level of genetic differentiation ($F_{ST} = 0.145$) detected between the two continuously distributed populations was similar to that previously measured ($F_{ST} = 0.150$) between samples from two sites 30 km apart. Here, because of our original sampling design, we can also argue that the genetic discontinuity appeared to be sharp rather than gradual, as we observed a break in allele frequency variation that clearly delineated two genetic entities. These two populations seem to be genetically homogeneous as no IBD signal and no mutation–migration–drift disequilibrium were detected within each

genetic entity. Hence, the fact that the IBD signal disappeared when considering both populations independently instead of analysing all samples appears to be an instance of the presence of spatially structured HWE populations (further exemplified in Guillot *et al.* 2009b; Fig. 1). However, the site located south of the genetic discontinuity presented higher levels of within-site diversity ($1-Q_0$) and could account for more subtle substructuring as evidenced by GENELAND results when considering the correlated allelic frequency model. These results might indicate that the area south of the genetic discontinuity has admixed genotypes from both populations.

We analysed the spatial genetic structure of our data set using different clustering algorithms, which did not always converge to the same clustering pattern. The spatially explicit method implemented in GENELAND detected several clusters, whereas the nonspatial software STRUCTURE failed to detect any genetic structure. Discrepancy among clustering methods has been reported frequently (Chen *et al.* 2007; Rowe & Beebe 2007; Frantz *et al.* 2009), but we found only one report in which no genetic structure ($K = 1$) was opposed to several clusters ($K > 1$) (Coulon *et al.* 2006). Such inconsistency can arise from differences in the underlying models, in the statistical estimators or in approximations in the algorithm used to compute this estimator, and it is generally difficult to disentangle the relative effects of these three sources of disagreement (Guillot *et al.* 2009b). It is important to keep in mind that the two programs used in this study are based on MCMC and are hence sensitive to convergence failures. This means that the outputs of these programs might, in certain cases, not be the exact solution of the mathematical equations but rather an approximation, the quality of which remains unknown. Here, it seems that STRUCTURE performs less well in detecting population structure than GENELAND.

Because we did not obtain fully congruent results from the different clustering methods (and in order to assess whether the population delineation is an artefact of the clustering method), we also used an interpolation procedure based on the $F/(1-F)$ classic genetic differentiation statistic. This method, which was developed for continuous population sampling, allowed us to confirm the main result obtained with GENELAND, i.e., an abrupt switch in gene frequency in the northeast part of the sampled domain, which clearly delineates two populations. Contrary to the clustering algorithms, this $F/(1-F)$ interpolation method is not prone to MCMC convergence issues. This result suggests that applying such an approach to different clustering analysis offers a convenient means of gathering solid information on spatial genetic population structures.

On the other hand, we found large similarities in GENELAND clustering results among the underlying genetic models. Interestingly, we did not find an effect of the spatial model, which attests that the genetic structure observed is not a by-product (e.g., edge effect) of considering spatial information in the clustering algorithm (same number of clusters and posterior probability membership values). Second, even if the number of clusters detected increases from two to four, when considering the correlated frequency model, a main discontinuity in gene frequency is still detected at the same place. Inconsistency between clustering concerning CFM and UFM in biological data has already been reported (Rosenberg *et al.* 2002) and illustrated by simulated data sets (Falush *et al.* 2003). The increase in the number of clusters detected ($K = 2$ under CFM vs. $K = 4$ under UFM) results from the CFM algorithm being more able to ascertain subtle structures (Guillot 2008) that potentially go undetected when using UFM. Accordingly, the genetic structure observed using CFM is simply a subpartition of $K = 2$ (detected by UFM) characterized by lower differentiation levels. This substructure could be attributable to the few sites with putative admixed genotypes (high values of $1-Q_0$).

Gene flow in natural populations can be influenced strongly by landscape features. For instance, two recent landscape genetics studies on animals (aphids and minks) (Lavandero *et al.* 2009; Zalewski *et al.* 2009) concluded that the presence of major rivers and mountains is barriers to gene flow. However, in the case of an airborne pathogen such as *M. fijiensis*, it is not easy to determine a priori which elements could represent an effective barrier to gene flow as no study to date has focused on this point (Archie *et al.* 2009). In the spatial domain sampled here, we considered three types of natural obstacle to *M. fijiensis* dispersal: a large river, forests and a high mountain. However, we found no correlation between genetic structure and any landscape feature or ecological variable. First, no obvious barrier matches the genetic discontinuity. Second, none of the hypothetical barriers occurring in the sampled domain (mountain, river, deep forests, etc.) appeared to have any effect on the population genetic structure of *M. fijiensis* (all obstacles to dispersal were included in the larger population). Such patterns, in which the landscape is not correlated with the genetic structure, have already been reported in other landscape genetics studies (Gauffre *et al.* 2008; Sahlsten *et al.* 2008). For instance, Sahlsten *et al.* (2008) explained the absence of geographical barriers between two genetically differentiated populations of hazel grouse in the Scandinavian Peninsula by the post-glacial reinvasion history of this species. We propose a similar argument. Indeed, we hypothesize that the present-day genetic structure

observed during this study could be better explained by the recent history of *M. fijiensis* expansion than by current gene flow. The earliest report of *M. fijiensis* in Africa was in 1973 in Zambia (Jones 2000). It arrived in Cameroon in 1980 from neighbouring Gabon and was restricted for a while to the southeastern area near the coast (Kribi region; Fouré & Lescot 1988) before being reported in 1983 close to the study area. North of this area, we could find no information on the spread of BLSD in Cameroon. The disease was then described for the first time in Nigeria in 1986. The pattern observed in this study, with the presence of two differentiated populations without any obvious barrier to gene flow, might reflect this history of northward expansion according to two conceivable scenarios. The first combines the effect of independent founder effects caused by the movement of a few individuals by long-distance spore dispersal (LDD or movement of infected material) as already suggested from analysis at the continental scale (Rivas *et al.* 2004), followed by natural gradual diffusion (GD). The genetic structure observed would then be attributable to secondary contact among two expanding patches. This original pattern has been recently empirically observed (termed 'stratified dispersal combination' by Parisod & Bonvin (2008) and supported theoretically by simulations (Nichols & Hewitt 1994; Ibrahim *et al.* 1996). However, we cannot reject a second likely scenario to explain the observed genetic structure. Indeed, recent studies have shown that low-frequency alleles can sometimes 'surf the wave' of the advance of a population range expansion, reaching high frequencies and spreading over large areas (Edmonds *et al.* 2004; Bialozyt *et al.* 2006; Excoffier & Ray 2008). This pattern, called 'gene surfing', has already been observed empirically in cultured bacterial populations (Hallatschek & Nelson 2008), although not yet in natural populations. In our study, the high variance of F_{ST} values detected among loci is congruent with the two hypotheses outlined above (stratified dispersal combination and gene surfing). Indeed, the drastic effects of genetic drift on genetic diversity during founder effects should randomly affect the different loci observed. Spatial genetic diversity patterns could theoretically help to better elucidate the origin of the observed genetic differentiation. Indeed, in both historical scenarios, we expect a decline in genetic diversity with increasing distance from the source population as a common pattern reflecting the recurrent founder effect influencing expanding populations (Hewitt 2000). Different genetic diversity patterns could result from these two scenarios, but these may not differ enough for statistical discrimination. Indeed, if the genetic differentiation we detected was caused by a secondary contact among two expanding patches, then two sources of

expansion would be present and we would expect a decrease in genetic diversity from the bottom to the top of the core of both populations. In the other case (gene surfing), we hypothesize the existence of only one source of expansion (in the south of the study area according to historical data), and in this case, a steady decrease in genetic diversity along the range of expansion (from south to north) would be predicted. However, in the present study, the number of sites in the northeastern population was too small to interpret the spatial distribution of the within-site diversity ($1-Q_0$) in this population and thus differentiate between the two scenarios. At a larger scale, the distribution of genotypes according to such patterns (stratified dispersal combination or gene surfing) could form a mosaic of relatively homogeneous patches (patchy genetic structure). Additional genetic structure analysis studies at a larger scale to test for such patterns would be of interest and could allow the origin of the observed genetic differentiation to be elucidated.

The extent of geographical variation in gene frequency is thought to result from a balance between forces tending to produce local genetic differentiation and forces tending to produce genetic homogeneity (Slatkin 1987). In our study, as discussed above, historical gene flow combined with genetic drift owing to small finite population size at the time of *M. fijiensis* introduction to Cameroon might be responsible for the current genetic structure observed (stratified dispersal combination or gene surfing). However, we assume high effective sizes in current *M. fijiensis* populations. Lesion-counting experiments on plantain leaves in Costa Rica revealed more than 15 000 lesions/leaf (Gauhl 1994). Even if not all lesions participate in the next generation, those that do might represent a large effective population size. Such a characteristic would have several consequences for both the evolution of *M. fijiensis* populations and the results we obtained. First, the absence of correlation between landscape features (a mountain, a deep forest and a river) and the genetic structure observed in this study should be interpreted taking into account the high effective sizes. Indeed, a recent simulation study showed that, for species with high effective size, recent barriers to effective dispersal can be difficult to detect through analysis of genetic variation (Gauffre *et al.* 2008). These landscape features could then act as effective barriers to contemporary gene flow although the high densities do not allow their detection by clustering methods. Second, the very high densities assumed might have prevented the detection of a significant IBD pattern on our specific transect. Indeed, according to the equation of the slope of the IBD regression line given by (Rousset 2000), ($b = \frac{1}{2N\pi\sigma^2}$), we can easily predict that high effective size (N) and/or

high dispersal abilities (σ^2) might be responsible for a weak IBD signal even with low dispersal abilities (σ^2). As a consequence, no inference on dispersal distance could be gained from the embedded transect using the IBD approach.

In summary, we have uncovered a very interesting situation, i.e., one in which two populations characterized by high effective size encounter each other. Our results could provide vital information related to the dispersal abilities of individuals. Indeed, if there is no differential selection between the two populations, each group will spread smoothly into the other's range and the genetic differentiation produced at the moment of contact will dissolve, leaving only a smooth, gradually falling cline (Endler 1977). If the population size is high enough to avoid the effects of genetic drift, then the rate of decay is thought to be affected only by quasi-deterministic spatial diffusion processes. Further analysis of the genetic discontinuity detected in the present study should be carried out after just a few more generations in order to test for the presence of such a falling cline pattern in an attempt to better characterize *M. fijiensis* dispersal processes.

In conclusion, our study has shown that individual-based population genetics analyses realized at a small scale can provide information on processes that have driven the recent colonization of an invasive plant pathogen. Indeed, using clustering analysis and a new F_{ST} -based procedure, we found evidence of a population structure in the *M. fijiensis* population in the southwestern plantain production area of Cameroon but found no evidence that geographical features affect current gene flow and dispersal patterns. We hypothesize that historical demographic events are responsible for the genetic structure detected and that high effective population sizes might have masked the contribution of contemporary processes to the genetic structure of these populations. This confusion between historical and contemporary processes has been poorly discussed in landscape genetic studies to date because the majority of such studies have focused on species characterized by moderate population sizes. However, plant pathogen populations are often characterized by recent expansion events and high densities. In this study, we showed that this latter demographic parameter should be taken into account more explicitly in future landscape genetics research.

Acknowledgements

We thank CARBAP (Oscar Nguidjo, Robert Dongmo and Josué Essoh Ngando) for technical and logistical support during sampling in Cameroon; Luc Pignolet for isolation of fungal strains; Arnaud Estoup, Renaud Vitalis, Thomas Lenormand, Etienne

Klein, Denis Bourguet, Benoit Barrès and Eric Bazin for helpful discussions; Dany Lo Seen for technical help in mapping georeferenced individuals; and Helen Rothnie for her attentive reading and careful English language checking. Data used in this work were produced partly through the molecular genetic analysis technical facilities of the IFR119 "Montpellier Environnement Biodiversité". Adrien Rieux was supported by a CIFRE doctoral fellowship from Bayer CropScience and the Agence Nationale de la Recherche et de la Technologie (ANRT-123/2008); and the Agence Nationale de la Recherche (ANR 07-BDIV-003/Emerfundis). This work was also supported by a Europaid contract ATF/UE- No 146-762/786/798/801.

References

- Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.
- Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution*, **24**, 21–30.
- Balkenhol N, Gugerli F, Cushman SA *et al.* (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology*, **24**, 455–463.
- Bialozyt R, Ziegenhagen B, Petit RJ (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, **19**, 12–20.
- Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, **96**, 523–536.
- Carlier J, Lebrun MH, Zapater MF, Dubois C, Mourichon X (1996) Genetic structure of the global population of banana black leaf streak fungus, *Mycosphaerella fijiensis*. *Molecular Ecology*, **5**, 499–510.
- Carlier J, De Waele D, Escalant JV (2002) *Global Evaluation of Musa Germplasm for Resistance to Fusarium Wilt, Mycosphaerella Leaf Spot Diseases and Nematodes* (ed. Vézina A, Picq C), INIBAP technical Guidelines 6. INIBAP, Montpellier.
- Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.
- Coulon A, Guillot G, Cosson JF *et al.* (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669–1679.
- Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology*, **17**, 2382–2396.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 975–979.
- Endler JA (1977) Geographic variation, speciation, and clines. *Monographs in Population Biology*, **10**, 1–246.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

- Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, **23**, 347–351.
- Fabre F, Plantegenest M, Mieuze L *et al.* (2005) Effects of climate and land use on the occurrence of viruliferous aphids and the epidemiology of barley yellow dwarf disease. *Agriculture Ecosystems & Environment*, **106**, 49–55.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- Fontaine MC, Baird SJE, Piry S *et al.* (2007) Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *BMC Biology*, **5**, 30.
- Fouré E, Lescot T (1988) Variabilité génétique des *Mycosphaerella* inféodés au genre *Musa*. Mise en évidence de la présence au Cameroun sur bananiers et plantains d'une cercosporiose (*Mycosphaerella musicola*) au comportement pathogène atypique. *Fruits*, **43**, 407–415.
- Francois O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics*, **174**, 805–816.
- Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.
- Garnier-Gere P, Dillmann C (1992) A computer program for testing pairwise linkage disequilibria in subdivided populations. *Journal of Heredity*, **83**, 239–239.
- Gaufré B, Estoup A, Bretagnolle V, Cosson JF (2008) Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology*, **17**, 4619–4629.
- Gauhl F (1994) *Epidemiology and Ecology of Black Sigatoka*, PhD Thesis. INIBAP, Montpellier, p. 120.
- Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.
- Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation—the correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222–2228.
- Guillot G, Santos F (2009) A computer program to simulate multilocus genotype data with spatially autocorrelated allele frequencies. *Molecular Ecology Resources*, **9**, 1112–1120.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.
- Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.
- Guillot G, Estoup A, Santos F (2009a) *Documentation for GENELAND Software: Version 3.1.5*. Available at: <http://www2.imm.dtu.dk/~gigu/Geneland/Geneland-Doc.pdf>.
- Guillot G, Leblois R, Coulon A, Frantz AC (2009b) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734–4756.
- Halkett F, Coste D, Rivas Platero GG *et al.* (2010) Isolation by distance and genetic discontinuities in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Molecular Ecology*, **19**, 3909–3923.
- Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theoretical Population Biology*, **73**, 158–170.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
- Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282–291.
- Jeger MJ (1989) *The Spatial Component of Plant Disease Epidemics*. Prentice-Hall Inc., Englewood Cliffs, NJ, 1–13.
- Jones DR (2000) *Diseases of Banana, Abaca and Enset*. CABI Publishing, Oxon.
- Kitron U (1998) Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *Journal of Medical Entomology*, **35**, 435–445.
- Lavandero B, Miranda M, Ramirez CC, Fuentes-Contreras E (2009) Landscape composition modulates population genetic structure of *Eriosoma lanigerum* (Hausmann) on *Malus domestica* Borkh in central Chile. *Bulletin of Entomological Research*, **99**, 97–105.
- Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a “Continuous” population under isolation by distance. *Molecular Biology and Evolution*, **20**, 491–502.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.
- Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions techniques with appropriate. *Trends in Ecology & Evolution*, **20**, 136–142.
- Marin DH, Romero RA, Guzman M, Sutton TB (2003) Black Sigatoka: an increasing threat to banana cultivation. *Plant Disease*, **87**, 208–222.
- Milgroom MG, Peever TL (2003) Population biology of plant pathogens – the synthesis of plant disease epidemiology and population genetics. *Plant Disease*, **87**, 608–617.
- Mourichon X, Fullerton RA (1990) Geographical distribution of the two species *Mycosphaerella musicola* Leach (*Cercospora musae*) and *M. fijiensis* Morelet (*C. fijiensis*), respectively agents of Sigatoka disease and Black Leaf Streak disease in bananas and plantains. *Fruits*, **45**, 213–218.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.
- Neu C, Kaemmer D, Kahl G, Fischer D, Weising K (1999) Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. *Molecular Ecology*, **8**, 523–525.
- Nichols RA, Hewitt GM (1994) The genetic consequences of long-distance dispersal during colonization. *Heredity*, **72**, 312–317.
- Parisod C, Bonvin G (2008) Fine-scale genetic structure and marginal processes in an expanding population of *Biscutella laevigata* L. (*Brassicaceae*). *Heredity*, **101**, 536–542.
- Parnell M, Burt PJA, Wilson K (1998) The influence of exposure to ultraviolet radiation in simulated sunlight on ascospores causing Black Sigatoka disease of banana and plantain. *International Journal of Biometeorology*, **42**, 22–27.
- Plantegenest M, Le May C, Fabre F (2007) Landscape epidemiology of plant diseases. *Journal of the Royal Society Interface*, **4**, 963–972.
- Pritchard JK, Wen W (2003) *Documentation for STRUCTURE Software: Version 2*. Available at: <http://pritch.bsd.uchicago.edu/software/structure2/readme.pdf>.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

14 A. RIEUX ET AL.

- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9197–9201.
- Rees EE, Pond BA, Cullingham CI *et al.* (2009) Landscape modelling spatial bottlenecks: implications for raccoon rabies disease spread. *Biology Letters*, **5**, 387–390.
- Rivas GG, Zapater MF, Abadie C, Carlier J (2004) Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis*. *Molecular Ecology*, **13**, 471–482.
- Robledo-Arnuncio JJ, Rousset F (2010) Isolation by distance in a continuous population under stochastic demographic fluctuations. *Journal of Evolutionary Biology*, **23**, 53–71.
- Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. In: *Plant Molecular Biology Manual* (eds Gelvin SB, Schilperoort RA), pp. A6, pp. 1–10. Kluwer Academic Publishers, Dordrecht.
- Rosenberg NA, Pritchard JK, Weber JL *et al.* (2002) Genetic structure of human populations. *Science*, **298**, 2381–2385.
- Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Rowe G, Beebe TJ (2007) Defining population boundaries: use of three Bayesian approaches with microsatellite data from British natterjack toads (*Bufo calamita*). *Molecular Ecology*, **16**, 785–796.
- Sahlsten J, Thorngren H, Hoglund J (2008) Inference of hazel grouse population structure using multilocus data: a landscape genetic approach. *Heredity*, **101**, 475–482.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.
- Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**, 479–498.
- Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
- Stover RH, Simmonds NW (1987) *Bananas*, Wiley & Sons, New York.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Zalewski A, Pierny SB, Zalewska H, Lambin X (2009) Landscape barriers reduce gene flow in an invasive carnivore: geographical and local genetic structure of American mink in Scotland. *Molecular Ecology*, **18**, 1601–1615.
- Zapater MF, Duchemin M, Dussart JF *et al.* (2008) Microsatellite markers for the fungal banana pathogens *Mycosphaerella fijiensis*, *Mycosphaerella musicola* and *Mycosphaerella eumusae*. *Molecular Ecology Resources*, **8**, 1121–1125.

This study was conducted by A.R. and F.H. during their PhD thesis and post-doctoral position in the team BEC of the research unit BGPI, respectively. The main objective of A.R.'s PhD thesis is to characterize *Mycosphaerella fijiensis* dispersal processes using different population genetics and epidemiological approaches. F.H. now works on the ecology and the population genetics of forest fungal pathogens, especially the poplar rust *Melampsora larici-populina*. M.-F.Z., L.L.B. and J.C. have a common research programme dealing mainly with the dynamics and evolution of populations of the fungus *Mycosphaerella fijiensis* pathogenic on banana. V.R. adapts population genetic methods and develops models to study the evolution of plant pathogenic fungi. One of F.R.'s research interests concerns estimation methods of dispersal rates under isolation by distance and statistical analysis of spatial genetic patterns.

No effect of recent range expansion and fungicide treatment on the spatial genetic structure of a plant pathogenic fungus in its agricultural landscape

A. RIEUX^{*}, L. DE LAPEYRE DE BELLAIRE[‡], V. RAVIGNE^{*} and J. CARLIER^{*}

^{*} CIRAD, UMR BGPI, Campus international de Baillarguet, TA A-54K, F-34398 [†]

[‡] CIRAD, Persyst, UPR Syst. Banan. Ananas, TA B-26 / PS4, Blvd. de la Lironde, 34398 [†]

[†] Montpellier Cedex 5, France

SUMMARY

Understanding how geographical and environmental features structure genetic variation at both the population and individual levels is crucial in biology, and especially in the case of pathogens. However, distinguishing between contemporary and historical event signatures on spatial genetic structure remains a challenging task. In the present study, we investigated the case of a plant pathogenic fungus that has recently colonized an agricultural landscape characterized by the presence of several types of potential barriers to gene flow and connectivity. First, we sampled around 300 isolates at a global scale of 50x80 Km on untreated plants. The study site included a river, low host densities areas and several industrial plantations treated with fungicides. However, using two different clustering algorithms we did not detect any genetic structure in the study area. Second, we specifically evaluated the hypothesis that, because of genetic drift, exposure to fungicides would reduce the genetic diversity of treated populations in comparison to untreated ones and influence the spatial genetic structure. At the local scale of an industrial plantation, we did not detect any significant reduction in neutral genetic diversity displayed by microsatellites markers in 6 treated plots in comparison to 6 untreated ones. At this local scale, genetic structure was assessed by analysing molecular variance. The AMOVA partition of the total genetic variance indicated that 99% of the variation occurred within populations. Our results suggest that the high effective size suspected for this pathogen fungus and/or a high level of gene flow between treated and untreated populations may be responsible for the absence of fungicide effect on genetic diversity and differentiation.

Key words: fungicides, genetic drift, landscape genetics, genetic diversity, agricultural landscape, *Mycosphaerella fijiensis*.

INTRODUCTION

Understanding how landscape elements and environmental factors influence spatial genetic structure of populations has important implications for evolutionary biology (Holderegger & Wagner 2008). In this research area, a discipline called landscape genetics, which combines approaches from landscape ecology, population genetics and spatial statistics, has emerged recently to specifically assess the contemporary effects of landscape composition on spatial genetic variation (Manel *et al.* 2003, Storfer *et al.* 2007). Within this context, the overwhelming majority of studies have addressed the question of how landscape features affect connectivity or create barriers to gene movement (see Sork *et al.* 2010, for a review); linear elements, such as rivers, mountain ridges or roads have been the most frequent barriers to gene flow tested to date. However, other less obvious ecological processes, for instance those linked to the presence of unsuitable natural habitat, might also influence population connectivity (Mc Rae & Beier 2007).

A major challenge for landscape genetic studies is the potential mismatch between the temporal and spatial scales of landscape and genetic data (Balkenhol 2009). Indeed, genetic data are an amalgamation of past and current processes. In such a context, historical expansion processes are known to be an important factor influencing the spatial genetic patterns seen at the present time (Excoffier *et al.* 2009). In addition to the presence of historical barriers to gene flow, natural dispersal abilities of species during range expansion can generate spatial breaks in allele frequencies. For instance, the combination of rare long-distance dispersal (LDD) events with natural gradual diffusion (GD) is expected to produce a patchy distribution of allele frequencies. This original pattern first gained theoretical support from simulations (Nichols & Hewitt 1994; Ibrahim *et al.* 1996) and more recently has been empirically observed and termed “stratified dispersal combination” by Parisod & Bonvin (2008). On the other hand, recent studies have shown that when dispersal is limited, low-frequency alleles can sometimes “surf the wave” of the advance of a population range expansion, reaching high frequencies and spreading over large areas (Edmonds *et al.* 2004; Bialozyt *et al.* 2006; Excoffier & Ray 2008). This pattern, called “gene surfing”, has already been observed empirically in cultured bacterial populations (Hallatschek & Nelson 2008), although not yet in natural populations. Because landscape and historical effects can be confounded, fine scale genetic structure analyses in two-dimensional environments could help to disentangle the relative influence of past and current processes on spatial genetic structure.

Plant pathogenic fungi are interesting models with which to study the effect of landscape elements and environmental factors on spatial genetic structure. The spread of such organisms is inherently a spatial process often embedded in heterogeneous landscapes (Biek & Real 2010). In addition to being fragmented by landscape elements, pathogen populations in modern agroecosystems are challenged regularly by a variety of cultural practices that aim to reduce plant infections (Stukenbrock and Mc Donald 2008). Among them, the intensive use of fungicides might have an important influence on agricultural landscape fragmentation and the connectivity of pathogen populations. This practice is the most common way of controlling fungi worldwide and represents a major ecological constraint for sensitive pathogen populations. Landscape genetics studies of pathogenic fungi might thus afford the opportunity to test the influence of the use of fungicides as

an obvious ecological element acting as a potential barrier to gene flow. However, colonization events combining LDD and range expansion are frequent in pathogenic fungi (Brown and Hovmøller, 2002) and could have also an impact on spatial genetic structure. It is thus of prime importance to distinguish contemporary and historical genetic signatures. Landscape genetics studies can provide opportunities to achieve this objective in the case of plant pathogenic fungi. Furthermore, because such organisms cause serious biological and economical damage in several agro-ecosystems (Anderson et al. 2004), a detailed understanding of how pathogen spatial genetic structure is influenced at the agricultural landscape scale is also likely to increase our ability to control plant disease (Archie et al. 2009). However, despite all these valuable advantages, landscape genetics approaches have thus far seldom been applied to plant pathogen populations (Biek & Real 2010).

Black leaf streak disease (BLSD) of banana is caused by the ascomycete fungus *Mycosphaerella fijiensis*. BLSD in South-West Cameroon banana-producing area constitutes an optimal testing ground for evaluating the relative influence of several factors on spatial genetic structure. BLSD is the most important foliar disease of this culture (Stover & Simmonds 1987) and intensive cultivation of dessert banana for export can be achieved only through the intensive use of fungicides (Marin et al. 2003, de Lapeyre de Bellaire *et al.* 2010). The agricultural landscape in Cameroon is characterized by the geographical proximity of both large industrial plantations treated with fungicides and many untreated fields. Such a situation is rare in traditional agricultural landscapes and provides an interesting opportunity to study the effect of chemical treatments on the connectivity of a pathogenic fungus. However, BLSD is also a recent pandemic. *M. fijiensis* spread from South East Asia to the Pacific Islands, Africa and Latin America in the 1970s (Carlier et al. 2000). *M. fijiensis* is a haploid, heterothallic fungus that spreads via three modes: the movement of infected plant material and the dispersal of spores produced via either sexual (ascospores) or asexual (conidia) reproduction (Gauhl 1994b). *M. fijiensis* arrived in Cameroon in 1980 from neighboring Gabon (Fouré & Lescot 1988). Isolation by distance (IBD) along a 300 km long transect in Cameroon suggested continuous range expansion at the regional scale through natural and gradual dispersal of spores (Halkett et al. 2010). However, a break in allelic frequencies was detected at the middle of the transect. A spatial analysis of population structure was recently conducted within a 30x40 Km area including this discontinuity (Rieux et al. 2011). This study allowed spatial mapping of a sharp genetic discontinuity but did not reveal any influence of landscape features, such as a high mountain, a natural forest (in which banana plants are practically nonexistent) or a broad river, on spatial genetic diversity. The genetic discontinuity was explained as a relict of the recent history of *M. fijiensis* expansion in the country rather than to have resulted from contemporary gene flow. However, because this first study was carried out in an area containing untreated populations only, the potential effect of fungicide treatments on spatial genetic structure could not be tested.

In the present study, we were mainly interested in understanding if a major ecological constraint, represented by the massive use of fungicides, can have an impact on the population connectivity of *M. fijiensis*. To achieve this objective, we analysed the spatial genetic structure of *M. fijiensis* populations within the so called “Moungo” production area in Cameroon. Several large treated

areas are distributed within this region and are surrounded by non-treated bananas plantations. Because the disease appeared recently (i.e., 1983) in this area and spread rapidly to all plantations of the region, analysis of the spatial genetic structure was also conducted to detect possible genetic patterns resulting from range expansion of *M. fijiensis*. This analysis was carried out with isolates collected following a two-dimensional sampling scheme and genotyped using microsatellite markers. To further test the potential effect of the presence of treated areas, the distribution of genetic diversity was analysed more precisely by comparing populations between two neighbouring treated and non-treated areas.

MATERIALS & METHODS

Study site and isolate sampling

All isolates were collected in South-West Cameroon in January 2009 according to two independent strategies. First, “2D” sampling was realised on a broader area of 50x80 km (i.e. 40.10⁶ ha) (see Figure 1). We collected 5 isolates on every 150 untreated plantain plants sampled. Each plant was sampled every 3.5 km along accessible roads of the study site. This area is situated just below the area sampled in 2006 (Rieux et al. 2011) and is “fragmented” by several potential natural barriers to gene flow, such as a river (width ~ 200 m), some areas characterized by lower host densities (because of industrial plantation of other crops) and three industrial treated bananas plantations. The sampled young lesions (stage 2, more details in Robert et al. 2010) were cut individually directly after sampling and dried for 1 day at 55°C before being stored at –20°C until analysis. Second, “treated vs. untreated” sampling was carried out inside and around a 3000 ha industrial banana plantation (Figure 1). We sampled 50 individuals (lesions) on every 12 banana plants. Six of these plants (T1–T6) were situated inside the industrial plantation (fungicide treated) and the other six (UT1–UT6) were outside the industrial plantation (fungicide untreated). Distances between each sampled plant and the closest border of the plantation were between 1 and 2 km.

Molecular analysis

Total genomic DNA was extracted directly from sampled lesions using the protocol described by Robert et al. (2010). Genotyping was carried out using 4 polymerase chain reaction (PCR) multiplex sets comprising 20 microsatellite loci (Neu et al. 1999, Zapater et al. 2008, Robert et al. 2010). PCR reaction conditions are also detailed in Robert et al. (2010). Among the 20 markers, 10 were tetra-nucleotide, 2 were tri-nucleotide and 8 were di-nucleotide microsatellite loci (See Table 1). Finally, the amplified fragments were separated on a 16 capillary ABI Prism 3130 XL sequencer (Applied Biosystems, Foster City, CA) and analysed for length variation using the GeneMapper® Software (Applied Biosystems). We genotyped 32 individuals/plant for the “treated vs untreated” sampling (32x12=384 individuals genotyped) and 2 individuals/plant for the “2D” sampling (2x150=300 individuals genotyped).

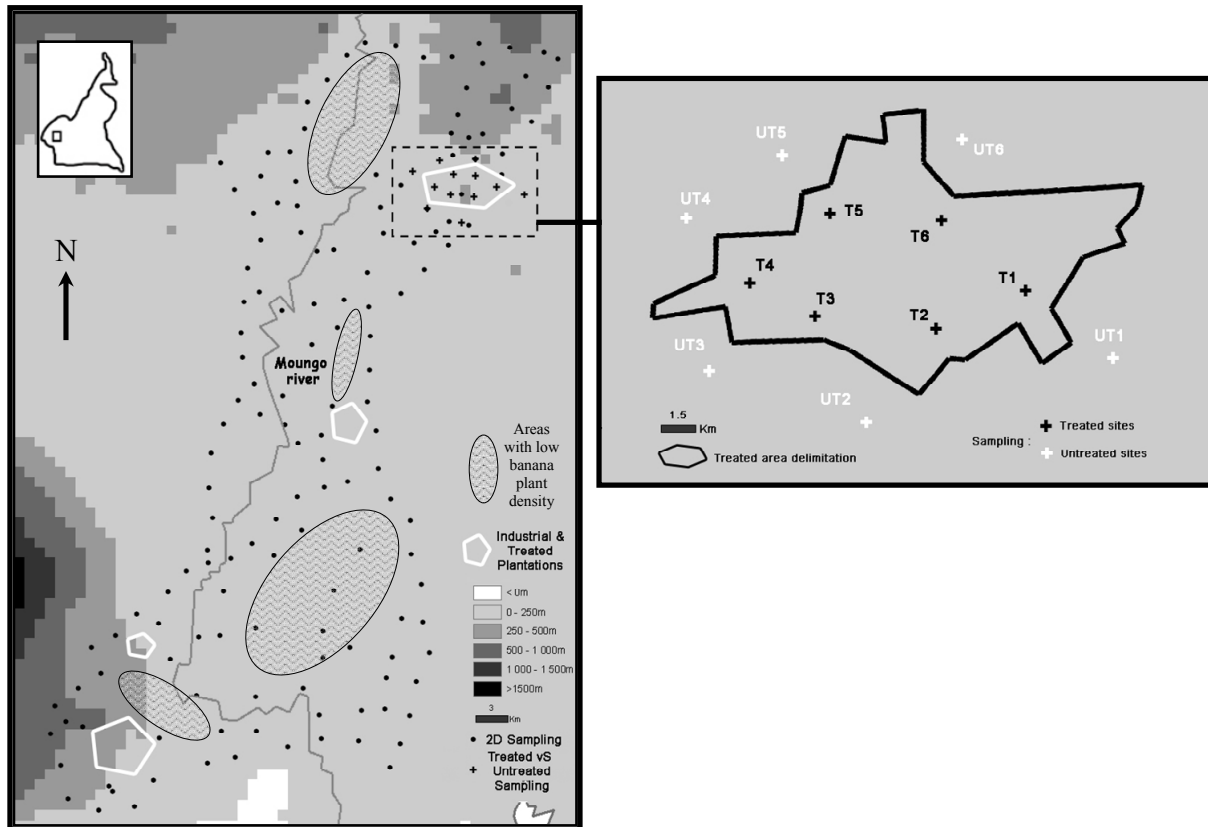


Figure 1. Representation of the 50x80 Km studied area localised in the region Centre of Cameroon and sampled sites schematization. *On the left:* Distribution of untreated sampled sites at a global scale (black dots) and of the “treated vs. untreated” sampling sites at a local scale (black crosses). *On the right:* Zoom on the “Treated vs Untreated” sampling area.

Analysis of genetic data

Basic population genetic analysis

Population gene diversity was estimated from the expected heterozygosity [H_E , unbiased estimate calculated following Nei (1978)] and the allelic richness (A_R). These two indices were calculated using FSTAT 2.9.3 (Goudet 1995). We used the non-parametric Wilcoxon-Mann-Whitney test to look for significant differences in heterozygosity, allelic richness and genotypic diversity between the treated and untreated populations. The genotypic diversity, defined as the probability that two individuals taken at random have different multilocus genotypes, was measured using MULTILOCUS (Agapow & Burt 2001). Linkage disequilibrium between all pairs of loci was tested using Fisher’s exact tests (Garnier-Gere & Dillman 1992) implemented in GENEPOP (Rousset 2008) run on the web [Markov chain parameters : dememorization number = 2000, number of batches = 250, number of iterations per batch = 2000]. We used the false discovery rate (FDR) procedure (Benjamin & Hochberg 1995) to control for multiple testing. The resulting adjusted P-values are called Q-values. This procedure is implemented in the R package QVALUE (Storey 2002). Because two individuals linked to a locus can increase their probability of being linked to another locus, we measured the multilocus linkage disequilibrium (\bar{r}_D) based on the index of association (Brown et al. 1980). This statistic is corrected for the number of loci. Departure from

random association of alleles across loci was assessed by bootstrapping alleles among individuals 1000 times independently for each locus using MULTILOCUS (Agapow & Burt 2001).

Analysis of population genetic structure

Clustering analysis. We carried out Bayesian population structure analysis on the “2D” sampling dataset using STRUCTURE v2.2 (Pritchard et al., 2000; Falush et al., 2003) and GENELAND v3.1.4 (Guillot et al., 2005). The base-unit of these methods are individual multilocus genotypes. Subdivision of the data into clusters in these Bayesian methods is done by minimizing Hardy-Weinberg disequilibrium and gametic phase disequilibrium between loci within groups. STRUCTURE bases its inference on genetic data only, whereas Geneland can take spatial localisation information for genotyped individuals explicitly into account. We followed the general recommendations of the authors of these methods in order to run the programs. Using STRUCTURE, we tested all values from $K=1$ to $K=10$ with a Markov chain MonteCarlo (MCMC) of 600,000 iterations following a burn-in of 60,000 iterations. Other parameters were set to default values (admixture model & correlated allele frequencies). For each K , we calculated the mean posterior probability over its runs and we used this value to estimate the posterior probability of each K using the formula given by Pritchard & Wen (2003). Five independent runs were performed to check for the consistency of the Markov chains convergence. GENELAND analyses were conducted using the correlated model of allele frequencies and long MCMC runs to ensure convergence of the chain (Guillot *et al.* 2009). We performed five independent runs with 1,000,000 iterations (thinning = 100) allowing K to vary from 1 to 10. The maximum rate of the Poisson process was fixed at 500, the maximum number of nuclei in the Poisson-Voronoi tessellation at 300, and the potential error for spatial coordinates at 2 m. K was inferred from the modal value of the run highest likelihood. IBD was assessed using an *ad hoc* haploid version of the “isolation by distance between individuals” suboption implemented in GENEPOP (version 4) (Rousset 2008, see Rieux et al. 2011 for more details on the haploid version). The ISOLDE procedure tests the correlation between genetic distances (a_r) and the logarithm of geographical distances (as suggested for 2-dimensional populations). Significance of regression slopes was evaluated from Mantel tests performed across 100,000 permutations. IBD was tested on the entire dataset.

AMOVA analysis. The existence of genetic structure between treated and untreated areas was tested using an analysis of molecular variance (AMOVA) (Excoffier et al, 1992, Michalakis and Excoffier, 1996) and a locus-by-locus AMOVA framework, as executed in Arlequin software. Total variance in gene frequencies was partitioned among groups (treated vs. untreated), among populations within groups and within populations by carrying out a hierarchical AMOVA on Euclidean pairwise distances among individuals. The significance of the fixation indices calculated at each level of analysis is tested using a non-parametric permutation approach described in Excoffier et al. (1992). Additionally, the level of genetic differentiation at microsatellite loci among populations and groups was estimated as F_{ST} (Weir and Cockerham 1984) using GENEPOP run on the web. We used Fisher’s exact test for population differentiation implemented in GENEPOP run on the web [Markov chain parameters: dememorization number = 2000, number of batches = 250,

number of iterations per batch = 2000] to test for significant values of F_{ST} between populations and groups.

RESULTS

Spatial genetic structure at the global scale

Genetic diversity and linkage disequilibrium

Gene diversity for each locus ranged from 0.007 to 0.608, with an average value of 0.35 and mean allelic richness ranging between 1.97 and 7.94 (3.20 on average) (Table 1). Genotypic diversity was very high (0.9999) (only 2 individuals had the same multilocus genotype). We observed low ($\bar{r}_D = 0.002$) and not significant values ($P = 0.211$) of multilocus estimates of linkage disequilibrium. No pairs of loci showed significant linkage disequilibrium after FDR control.

Spatial genetic structure

The results of the Bayesian population structure analyses did not show any significant genetic differentiation across the study area. First, the estimated logarithm of likelihood for data analysed with STRUCTURE was highest for $K=1$, and the posterior probability of having one population was equal to 1 (Table 4). For $K > 1$, none of the 300 individuals could be assigned to a given cluster as each individual had a similar probability of belonging to each cluster. In GENELAND, the mean posterior density distribution (in %) of the number of clusters estimated over the 5 replicates was higher for $k=1$. Second, we did not find any significant correlation between geographical and genetic distances (slope of the regression $b=0.005$, confidence interval $\{-0.008;0.018\}$, P value = 0.267).

K	STRUCTURE		GENELAND
	mean $\ln P(D)$	$\Pr(K=)$	% runs (K)
1	-3712.5	1	58.48
2	-3908.9	2.0E-113	15.75
3	-3971.9	7.6E-206	9.28
4	-4184.8	9.5E-238	4.04
5	-5085.7	0	1.71
6	-5742.5	0	0.49
7	-6595.7	0	0.19
8	-5215.3	0	0.06
9	-4864	0	0
10	-5000	0	9.91

Table 4. Results of individual-based Bayesian clustering analyses computed from $K=1$ to $K=10$. STRUCTURE analyses: mean values of likelihood over the 10 repetitions and posterior probability of having K clusters. GENELAND analyses: mean posterior density distribution (in %) of the number of clusters estimated over the 5 replicates.

Comparison between neighbouring treated and non-treated populations

Genetic diversity and linkage disequilibrium (LD)

Gene diversity for each locus ranged from 0.005 (425) to 0.555 (Fe05b), with an average value of 0.347, and mean allelic richness ranged between 2 and 7 (3.54 on average) (Table 1). Genotypic diversity was very high (0.9999), which means that very few individuals (only 3 out of 384) had the same multilocus genotype. We observed low ($\bar{r}_D = 0.001$) and not significant values ($P = 0.116$) of multilocus estimates of LD. When comparing mean levels of genetic diversity over the 20 loci between treated and untreated populations, we found no significant different differences in the values of both heterozygosity and allelic richness between treated and untreated populations (Table 2). Genotypic diversity was very high in both treated (0.9998) and untreated (0.9999) plots so that no significant differences between these populations was detected ($P=0.987$). We observed low and not significant values of multilocus estimates of LD in both treated and untreated populations (not shown). When testing for LD between pairs of loci independently on each population, only two pairs of loci showed significant LD and this signal was detected solely in the two treated populations: Fe09&401 in population T5 and 401&407 in population T2.

Locus	"2D" data game (N=300)		"Treated vs Untreated" data game (N=378)	
	He	Ar	He	Ar
401 ⁽⁴⁾	0.244	2.97	0.345	4.00
403 ⁽⁴⁾	0.054	6.89	0.067	6.99
405 ⁽⁴⁾	0.494	2.00	0.484	3.00
407 ⁽⁴⁾	0.48	2.00	0.493	5.00
413 ⁽⁴⁾	0.452	2.00	0.456	3.00
417 ⁽⁴⁾	0.477	2.00	0.498	3.00
424 ⁽⁴⁾	0.5	2.97	0.494	2.00
425 ⁽⁴⁾	0.007	1.97	0.005	2.00
428 ⁽⁴⁾	0.388	2.00	0.331	3.00
434 ⁽⁴⁾	0.493	2.00	0.49	2.99
F26b ⁽²⁾	0.13	3.96	0.056	3.00
F40 ⁽²⁾	0.458	2.98	0.423	2.00
F5bis ⁽²⁾	0.465	2.00	0.464	2.00
FE05B ⁽²⁾	0.608	3.00	0.555	3.00
FE09 ⁽²⁾	0.44	4.94	0.304	5.99
Ff09 ⁽²⁾	0.11	3.97	0.031	3.00
N061 ⁽²⁾	0.026	2.97	0.01	2.99
N137 ⁽²⁾	0.517	7.94	0.432	7.00
N194b ⁽³⁾	0.363	3.00	0.456	3.99
N203 ⁽³⁾	0.565	3.97	0.539	3.00
All loci	0.364	3.276	0.347	3.547

Table 1. Summary of basic information and global population genetics analyses of "2D" and "Treated vs. Untreated" data sets. N, sample size; H_e , unbiased estimate of gene diversity; A_r , allelic richness are computed among the different loci. The number of nucleotides corresponding to the repet unit for each locus is indicated inside the brackets.

Locus	He			Ar			F_{ST}
	N=186	N=192		N=186	N=192		
	Untreated	Treated		Untreated	Treated		
401	0.293	0.394	NS	2.00	3.00	**	0.010 **
403	0.063	0.071	NS	2.00	2.17	NS	-0.004 NS
405	0.466	0.499	NS	2.00	2.17	NS	0.005 NS
407	0.48	0.503	NS	2.00	3.17	**	0.005 **
413	0.46	0.454	NS	2.00	2.17	NS	-0.005 NS
417	0.5	0.498	NS	2.00	2.17	NS	-0.004 NS
424	0.496	0.494	NS	2.00	2.00	NS	-0.005 NS
425	0	0.01	NS	1.00	1.17	NS	0.000 NS
428	0.311	0.351	NS	2.17	2.00	NS	-0.001 NS
434	0.484	0.497	NS	2.00	2.17	NS	-0.004 NS
F26b	0.041	0.07	NS	1.67	1.67	NS	-0.002 NS
F40	0.454	0.385	NS	2.00	2.00	NS	0.012 NS
F5bis	0.475	0.453	NS	2.00	2.00	NS	-0.002 NS
FE05B	0.545	0.568	NS	2.83	3.00	NS	-0.004 NS
Fe09	0.298	0.309	NS	2.17	3.17	*	0.002 *
ff09	0.021	0.041	NS	1.33	1.50	NS	0.001 NS
N061	0.011	0.01	NS	1.17	1.17	NS	-0.003 NS
N137	0.456	0.411	NS	4.50	4.67	NS	-0.002 NS
N194bB	0.431	0.479	NS	2.33	2.50	NS	0.001 NS
N203	0.549	0.53	NS	2.67	2.83	NS	-0.003 NS
All loci	0.342	0.351	NS	2.09	2.33	NS	0.0001 NS

Table 2: Summary of genetic diversity indices calculated on treated and untreated populations independently. H_e , unbiased estimate of gene diversity; A_r , allelic richness and significant differences are computed among the different loci. Pairwise genetic distances (F_{ST}) between treated vs. untreated populations are also indicated.
(NS $P>0.05$ * $P<0.05$; ** $P<0.01$).

Population genetic structure

The AMOVA indicated that most of the total genetic variation (99%; $P = 0.027$) was within populations (Table 3) but the analysis did not indicate any significant effect of either “populations within group” (0.21%; $P = 0.237$) or “group” (0.78%; $P = 0.185$). F_{ST} values calculated with GENEPOP were consistent with AMOVA results. Indeed, multilocus values of F_{ST} were weak and not significant between groups (treated populations vs untreated populations) and between each population (Table 2).

Source of variation	df	Sum of squares	Variance components	Percentage of variation	Fixation index (F)	P -value
Among groups	1	8.969	0.02734 V_a	0.78	0.00017	0.18530
Among populations within groups	10	36.933	0.00744 V_b	0.21	0.00215	0.23752
Within populations	374	1291.844	3.45413 V_c	99	0.00997	0.02700

Table 3. Hierarchical AMOVA based on nuclear microsatellite DNA: percentage of variation explained by different source of variation, variance components and F statistics.

DISCUSSION

In a previous study, Rieux *et al.* (2011) detected a sharp genetic discontinuity in the spatial genetic structure of *Mycosphaerella fijiensis* over a 30x40 km area fragmented by several potential physical barriers to gene flow, i.e. a mountain, a river and a deep forest. However, since the position of the genetic discontinuity was not correlated with any detectable barrier to gene flow, the authors hypothesized that a historical colonization event might be at the origin of the spatial break in allelic frequencies. This result suggested that the historical signal caused by *M. fijiensis* range expansion can persist and still be detectable nowadays. In the present study, we investigated the spatial genetic structure of this plant pathogenic fungus over a wider area (50x80 km) localized just south of the area analysed in the previous study. In addition to some landscape elements, this new area also includes the presence of a large plantation treated with fungicides, which could have an impact on pathogen populations. To our knowledge, this is the first landscape genetics study in which the potential effect of chemicals on gene flow has been considered. From a landscape genetics point of view, most previously published studies have focused on the influence of physical elements on gene flow (Sork *et al.* 2010), whereas less obvious ecological processes, linked to the presence of unsuitable natural habitat for instance have not been considered to any great extent as yet. One of the reasons for this may be the difficulty of mapping such an “ecological” barrier to gene flow since these are rarely as sharp and well-defined spatially as the case of pesticide use.

A single homogeneous population following a recent range expansion

Using a landscape genetic approach over the whole study area, we did not detect any spatial genetic structure in the *M. fijiensis* population. Such an absence of spatial genetic structure pattern can be informative regarding the dispersal processes that might have affected the spread of *M. fijiensis* over the studied area. The pathogen was first observed in 1983 in the Moungo producing area (Fourre & Lescot 1988), but there is no precise historical data indicating the spatial origin of this introduction and the direction of *M. fijiensis* expansion. Our result indicates that colonization of the area probably arose from a single introduction. Moreover, theoretical studies have shown that a spatial genetic structure can be generated following species range expansion, either by stratified dispersal combination or by gene surfing (Excoffier et al. 2009). Gene surfing has been shown to be favoured by small and rapidly growing populations displaying limited dispersal (Klopfstein et al. 2006). On the other hand, a stratified dispersal combination is expected to generate patchy genetic structure in the presence of rare LDD events (Bialozyt et al. 2006). But when LDD events are more frequent, it has been shown recently, and theoretically, that gene mixing is favoured, decreasing the probability of observing a spatial break in allelic frequencies through gene surfing or stratified dispersal combination (Excoffier et al. 2009, Fayard et al. 2009). Following these theoretical considerations, the absence of structure observed in the present study might suggest that *M. fijiensis* colonized the Moungo area through the expansion of large populations and frequent LDD events. Such a hypothesis is consistent with the biology of *M. fijiensis* since both large population size (Rieux et al. 2011) and frequent long-distance ascospore movement by wind dispersal (Burt et al. 1998, Parnell et al. 1998) have been suspected previously.

No effect of landscape or environmental elements on genetic structure

The landscape and environmental elements present in the Moungo area had no effect on the spatial genetic structure of *M. fijiensis* populations. Three kinds of elements in the agricultural landscape analysed could have a potential impact on *M. fijiensis* dispersal and connectivity. The first is a “physical” barrier represented by a river that is approximately 200 m wide. The second is a “biological” barrier represented by some wide areas in which host density is particularly weak. Such areas might decrease the connectivity between individuals at the global scale. Finally, the third element is a “chemical” barrier represented by the presence of four industrial banana plantations sprayed with fungicide and distributed within the studied area. These chemical barriers were already present at the time of *M. fijiensis* colonization because another related pathogenic fungus (*Mycosphaerella musicola*) was implanted in the area (Fourré & Lescot 1988). Treated areas are thought to constrain gene movement because of the weak probability of an individual being able to reproduce and disperse when inside such areas. To further test the potential effect of fungicide treatment on *M. fijiensis* genetic diversity, we conducted a complementary analysis at the scale of one particular banana industrial plantation. Here also, we found no difference in neutral genetic diversity (as evidenced by the expected heterozygosity, the number of alleles and the genotypic diversity measured at 20 microsatellite loci) between population from treated and untreated banana plants. Using both an AMOVA and Pairwise F_{ST} comparisons, we also detected no significant influence of fungicide treatment on genetic structure.

*Does the absence of spatial genetic structure mean that landscape and environmental elements haven't acted as barriers to gene flow in *M. fijiensis* populations?*

We hypothesized that the previously suspected high aerial dispersal abilities of *M. fijiensis* can easily explain why the river did not impact gene flow during its historical expansion. Moreover, because untreated plantain is distributed all around “host low densities” areas and industrial plantations, the expansion wave could have easily bypassed such elements. Thus, these elements probably did not constraint the spread of *M. fijiensis* during its expansion over the study area. However, once populations have been established in an area, these elements could have played a role as barriers to gene flow.

In this study, we particularly focused on the potential effect of the intensive use of fungicide in industrial plantations. A decrease in genetic diversity in treated areas in comparison to untreated ones has been shown previously in other pathogenic fungi (Grünwald et al. 2006, Matasci et al. 2008, Dobrowolski et al. 2008). But these populations were clonal and directional selection at some loci certainly influenced the whole genome through selective sweep (Fisher 1930). By contrast, the populations of *M. fijiensis* are random mating at the considered scale and present a low level of genetic disequilibrium due to recombination. Thus, the probability of detecting a direct effect of selection on 20 genetic markers chosen randomly and a priori neutral is very low. However, random changes in allele frequency can occur on the whole genome by genetic drift when effective population size decreases due to application of *chemicals*. Furthermore, for several years, protectant fungicides have been applied almost weekly inside the treated plantations (de Lapeyre et al. 2010). Such applications enable the control of the disease by preventing penetration of the pathogen into the leaves through a multisite mode of action. The airplane-delivered fungicide spray does not cover 100% of the surface of the leaves, resulting in the existence of a surviving population. But, such a drastic decrease in density also limits the occurrence of sexual reproduction. Indeed, *M. fijiensis* is self-incompatible (haploid heterothallic) so that the production of ascospores requires large enough population sizes for two sexually compatible lesions to spread and collapse on the same banana leaf. As a result, reproduction inside the treated area is realised mainly through the production of asexual spores (i.e., conidia), which are thought to disperse over very short distances (in contrast to sexual spores, i.e., ascospores) (Gauhl 1994).

The absence of neutral diversity difference between treated and untreated populations of *M. fijiensis* can be explained either because i) the remaining population size inside the treated area remains large enough to avoid the occurrence of genetic drift, and/or ii) a large amount of migration from untreated to treated populations continuously restores the genetic diversity inside the plantation. Thus, we hypothesized that the suspected high effective population size of *M. fijiensis* and/or high level gene flow may be responsible for the absence of any genetic drift effect on neutral diversity. However, data from the present study did not allow us to distinguish between these two hypotheses and it is very likely that both processes impede the occurrence of genetic drift simultaneously. The same arguments are also valid for the other elements present in the studied area. In other words, the landscape features considered as hypothetical physical, biological or chemical obstacles in the

present study could effectively act as effective barriers to contemporary gene flow but the high population densities do not allow detection of their effect using population genetics methods.

General conclusion

In summary, we have shown in the present study that historical expansion of *M. fijiensis* in the Moungo producing area has resulted in the installation of one single homogeneous population. This result suggests that that colonization of the area probably arose from a single introduction. However, we did not find any influence of the intensive use of fungicides on neutral genetic diversity and spatial genetic structure. We hypothesized that high effective sizes and/or high levels of gene flow may have prevented the detection of such hypothetical contemporary genetic drift influence on genetic diversity and structure. Characteristics as high effective size in the specific case of species that followed a recent expansion, as suspected for *M. fijiensis*, could constitute a limit to detect the effects of contemporary factors from population- and landscape genetics studies. However, this approach remains interesting as a tool with which to better reconstruct and understand the patterns of historical invasion and expansion for such species.

REFERENCES

- Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* **1**, 101-102.
- Anderson PK, Cunningham AA, Patel NG, *et al.* (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**, 535-544.
- Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution* **24**, 21-30.
- Balkenhol N, Gugerli F, Cushman SA, *et al.* (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* **24**, 455-463.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300.
- Bialozyt R, Ziegenhagen B, Petit RJ (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology* **19**, 12-20.
- Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology* **19**, 3515-3531.
- Brown JKM, Hovmoller MS (2002) Epidemiology - Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537-541.
- Burt PJA, Rutter J, Ramirez F (1998) Airborne spore loads and mesoscale dispersal of the fungal pathogens causing Sigatoka diseases in banana and plantain. *Aerobiologia* **14**, 209-214.
- Carlier J (2000) Population genetic structure and dispersal of *Mycosphaerella fijiensis*. *Info musa* **13**, 17-20.
- De Lapeyre de Bellaire L., Fouré E., Abadie C., J. C (2010) Black leaf streak disease is challenging the banana industry. *Fruits* **65**, 327-342.
- Dobrowolski MP, Shearer BL, Colquhoun IJ, O'Brien PA, Hardy GES (2008) Selection for decreased sensitivity to phosphite in *Phytophthora cinnamomi* with prolonged use of fungicide. *Plant Pathology* **57**, 928-936.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 975-979.
- Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology Evolution and Systematics* **40**, 481-501.
- Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution* **23**, 347-351.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
- Fayard J, Klein EK, Lefevre F (2009) Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology* **22**, 2171-2182.

- Fisher RA (1930) *The Genetical Theory of Natural Selection*
- Gauhl F (1994) Epidemiology and ecology of black sigatoka (PhD Thesis), p. 120. INIBAP, Montpellier.
- Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
- Grunwald NJ, Sturbaum AK, Montes GR, *et al.* (2006) Selection for fungicide resistance within a growing season in field populations of *Phytophthora infestans* at the center of origin. *Phytopathology* **96**, 1397-1403.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.
- Guillot G, Estoup A, Santos F (2009) Documentation for GENELAND software: Version 3.1.5.
- Halkett F, Coste D, Rivas Platero GG, *et al.* (2010) Isolation by distance and genetic discontinuities in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Molecular Ecology*.
- Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theoretical Population Biology* **73**, 158-170.
- Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* **58**, 199-207.
- Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**, 282-291.
- Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* **23**, 482-490.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.
- Marin DH, Romero RA, Guzman M, Sutton TB (2003) Black sigatoka: An increasing threat to banana cultivation. *Plant Disease* **87**, 208-222.
- Matasci CL, Gobbin D, Scharer HJ, Tamm L, Gessler C (2008) Selection for fungicide resistance throughout a growing season in populations of *Plasmopara viticola*. *European Journal of Plant Pathology* **120**, 79-83.
- McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19885-19890.
- Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**, 1061-1064.
- Neu C, Kaemmer D, Kahl G, Fischer D, Weising K (1999) Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. *Molecular Ecology* **8**, 523-525.
- Nichols RA, Hewitt GM (1994) The genetic consequences of long-distance dispersal during colonization. *Heredity* **72**, 312-317.
- Parisod C, Bonvin G (2008) Fine-scale genetic structure and marginal processes in an expanding population of *Biscutella laevigata* L. (Brassicaceae). *Heredity* **101**, 536-542.
- Parnell M, Burt PJA, Wilson K (1998) The influence of exposure to ultraviolet radiation in simulated sunlight on ascospores causing Black Sigatoka disease of banana and plantain. *International Journal of Biometeorology* **42**, 22-27.

- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Rieux A, Halkett F, de Bellaire LD, *et al.* (2011) Inferences on pathogenic fungus population structures from microsatellite data: new insights from spatial genetics approaches. *Molecular Ecology* **20**, 1661-1674.
- Robert S, Rieux A, Argout X, Carlier J, Zapater MF (2010) Optimized genotyping with microsatellite markers in the fungal banana pathogen *Mycosphaerella Fijiensis*. *American Journal of Botany* **97**, E130-E132.
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- Sork VL, Waits L (2010) Contributions of landscape genetics - approaches, insights, and future potential INTRODUCTION. *Molecular Ecology* **19**, 3489-3495.
- Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**, 479-498.
- Storfer A, Murphy MA, Evans JS, *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128-142.
- Stover RH, Simmonds NW (1987) Bananas. Willey & Sons, New York.
- Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology* **46**, 75-100.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
- Zapater MF, Duchemin M, Dussart JF, *et al.* (2008) Microsatellite markers for the fungal banana pathogens *Mycosphaerella fijiensis*, *Mycosphaerella musicola* and *Mycosphaerella eumusae*. *Molecular Ecology Resources* **8**, 1121-1125.

Annexe 1: Estimation de taille efficace chez une population Camerounaise de *Mycosphaerella fijiensis*

Analyses réalisées en collaboration avec Renaud Vitalis (CBGP-Montpellier)

Présentation de la démarche

On appelle effectif génétique (N_E) ou encore taille efficace, le nombre d'individus que comporterait une population dite « idéale » (i.e., constituée d'un effectif constant d'individus se reproduisant en panmixie, dans laquelle la dérive serait la seule force évolutive et dont la génération suivante serait constituée à partir de tirages aléatoires dans une urne infinie) qui montrerait les mêmes propriétés de dérive que la population étudiée. Cet effectif représente un paramètre très important en génétique des populations du fait de son influence directe sur l'intensité de la dérive génétique, dérive qui par interaction avec les autres forces évolutives détermine l'intensité et la distribution de la variabilité génétique au sein des populations naturelles. Ainsi, différentes méthodes indirectes ont été développées dans l'objectif de pouvoir quantifier ce paramètre critique à partir de données génétiques (voir pour revue Wang 2005). Dans ce domaine, il existe 3 grandes classes d'approches basées sur la mesure i) d'un excès d'hétérozygotes dans la descendance (Luikart & Cornuet 1999), ii) du déséquilibre de liaison (Hill 1981) ou bien iii) de la variation temporelle des fréquences alléliques (Nei and Tajima 1981; Waples 1989; Williamson and Slatkin 1999). Cependant, toutes ces méthodes font l'hypothèse d'équilibre génétique et démographique.

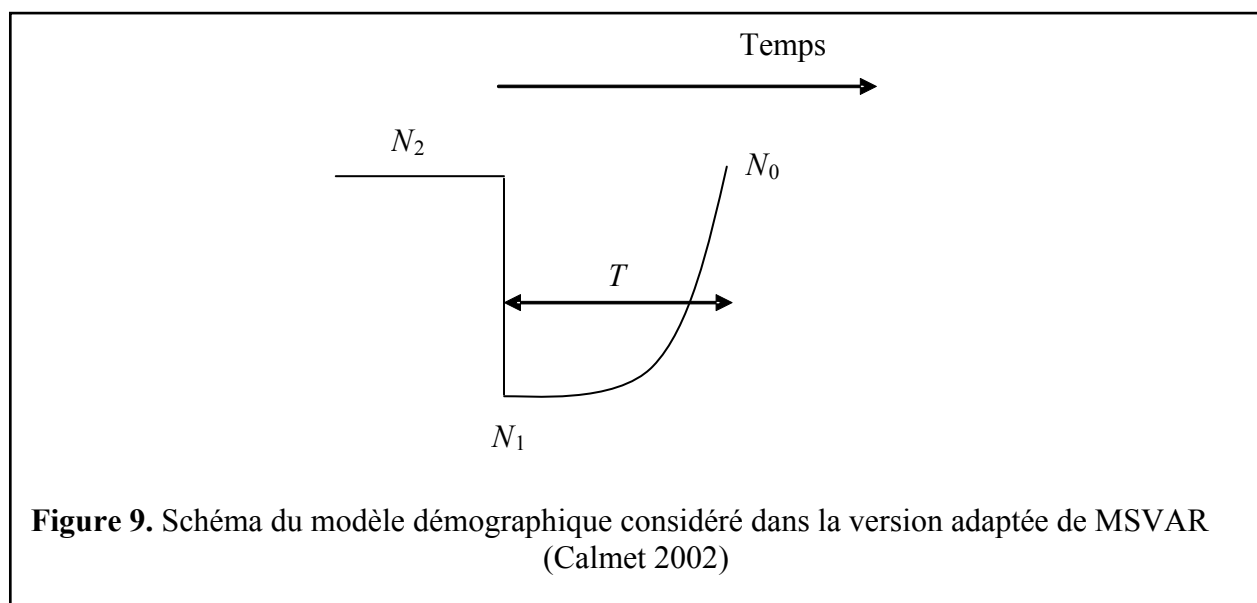
Nous avons cherché à obtenir une estimation de la taille efficace d'une population de *Mycosphaerella fijiensis* au Cameroun. Cependant, aussi bien l'histoire du pathogène au sein du pays que les niveaux de diversité génétiques semblent suggérer que les populations actuelles du pathogène ne sont pas à l'équilibre génétique. En effet, lors de l'arrivée du champignon au début des années 80, les populations fondatrices ont de toute évidence subi un fort goulet d'étranglement (suite à l'introduction initiale suspectée d'un faible nombre d'individus). Cet effet de fondation ayant entraîné une perte drastique de diversité génétique en comparaison avec les populations ancestrales asiatiques (Rivas et al. 2004). Par la suite, l'expansion spatiale du pathogène a été accompagnée d'une rapide augmentation des tailles « physiques » de population du fait de la biologie du pathogène (beaucoup de spores produites) et du bananier (procurant une réceptivité constante au pathogène tout au long de l'année). Cependant, la restauration de la diversité génétique est plus lente que la croissance démographique des populations (Maruyama & Fuerst 1985, Boileau et al. 1992). Elle l'est d'ailleurs d'autant plus que la taille physique des populations est grande et le fait que les niveaux de diversité soient actuellement toujours inférieurs aux niveaux asiatiques peut laisser penser que l'équilibre génétique n'a toujours pas été atteint. En conséquence, il n'était pas possible d'utiliser les méthodes indirectes classiques d'estimation de taille efficace dans le cas particulier de *M. fijiensis*. Nous avons donc cherché à utiliser une approche considérant de façon explicite de telles variations démographiques passées. Plusieurs méthodes ont récemment été développées dans ce domaine (e.g., Wilson and Balding 1998, Beaumont 1999, Drummond and Rambaut 2007, Hey and Nielsen 2004).

Présentation du modèle considéré

Parmi elles, l'approche MSVAR initialement développée par Beaumont (1999) puis améliorée par Storz et Beaumont (2002) a pour but la détection d'évènements démographiques passés ainsi que l'estimation des paramètres démographiques caractérisant ces évènements. Cette méthode a reçu une attention particulière ces dernières années (Girod et al. 2011). Le modèle démographique sous-jacent initial à cet algorithme considère une population isolée des flux géniques et ayant subi un changement de taille efficace dans le passé à la suite d'un évènement d'expansion ou de déclin. Ce modèle démographique a par la suite été adapté au cas particulier de fondations suivies d'explosion démographique (Calmet 2002) (Figure 9) qui semble bien correspondre à la situation de *M. fijiensis* au Cameroun. Nous avons appliqué cette méthode sur un jeu de données tronqué issu de l'échantillonnage présenté dans l'article n°1. La troncation a consisté à supprimer du jeu de données tous les individus appartenant à la petite population détectée au nord-ouest du domaine d'étude ainsi que les individus potentiellement issus d'admixture au sein de la zone de fracture. Par sécurité, tous les individus de l'axe Nord-Sud de l'article n°3 ont été supprimés du jeu de données afin de se rapprocher au maximum d'une situation caractérisée par la présence d'une seule population isolée des flux géniques. Un total de 398 individus génotypés à 12 marqueurs microsatellites a été conservé dans le jeu de données analysé avec MSVAR.

a) Le modèle démographique

Dans ce modèle adapté, on considère une population close, panmictique, de N_t génomes haploïdes. La taille de la population à la date t_0 ou cette population est échantillonnée est notée N_0 . On considère qu'à un moment t_2 dans le passé, une population de taille N_2 gènes jusque-là constante tombe brutalement à N_1 gènes (fondation) avant de croître immédiatement et de façon exponentielle pour atteindre la taille de population échantillonnée (explosion).



b) Le modèle génétique

Cette approche considère des marqueurs de type microsatellites. Le modèle génétique correspond donc au modèle mutationnel lié à ces marqueurs. Il s'agit d'un modèle dit à deux phases (TPM, Dirienzo et al. 1994) dans lequel, sachant qu'une mutation se produit, elle a une probabilité p d'être de type « pas » (perte ou gain d'une unité de répétition) et $(1 - p)$ d'être de type « saut », d'amplitude définie par une loi géométrique de moyenne $j \geq 1$.

c) Les paramètres

Les paramètres estimés du modèle sont :

- i) des paramètres de tailles efficace N_t : N_0 , N_1 et N_2
- ii) un paramètre de temps T depuis l'évènement de fondation
- iii) un paramètre μ pour le taux de mutation

Ces paramètres sont dits « naturels » pour les distinguer des paramètres dits « mis à l'échelle » également estimés par le modèle :

- i) le temps mis à l'échelle de la taille actuelle de population (T / N_0)
- ii) un paramètre classique de la coalescence ($\theta = 2 N_t \mu$) estimé pour chacune des tailles

d) Paramétrage du modèle et lancement des estimations

MSVAR repose sur une chaîne de Markov par simulation de Monte Carlo (MCMC) qui utilise l'algorithme de Metropolis-Hastings (Hastings 1970) pour estimer les distributions postérieures des paramètres d'intérêt du modèle. Le modèle autorise la variation des paramètres démographiques et mutationnels entre locus. La distribution des priors pour ces paramètres est sous le contrôle d'hyper-priors. D'une manière générale, nous avons choisi de considérer des distributions de priors très peu informatives pour les tailles de populations (car nous n'avons pas d'information). Nous avons utilisé les données historiques relatant l'arrivée de *M. fijiensis* en 1980 au Cameroun (Fouré & Lescot 1988) pour intégrer un prior sur le temps depuis l'évènement centré sur 250 générations mais nous avons tout de même conservé une grande variance pour que d'autres valeurs puissent être explorées. Pour le taux de mutation, nous avons choisi une distribution relativement peu informative centrée sur 10^{-4} .

Cinq analyses indépendantes ont été réalisées, avec des valeurs de départ de la chaîne de Markov différentes pour le générateur de nombres aléatoires initiant la séquence. Ces 5 analyses indépendantes permettent d'explorer l'espace des paramètres en partant de différents points et selon différentes séquences d'états. Cela permet de vérifier la convergence des chaînes : si toutes les chaînes aboutissent à la même distribution postérieure des paramètres, cela signifie qu'elles ont bien convergé. Si en revanche elles n'aboutissent pas à la même distribution, cela signifiera qu'elles n'ont pas atteint le même pic de vraisemblance et donc qu'elles n'ont pas convergé. Chacune de ces chaînes a été lancée initialement pour 10^6 itérations. Les premiers 10% de chaque chaîne ont été retirés en tant que préchauffage afin de minimiser l'effet du choix des valeurs de départ des paramètres sur les distributions postérieures des paramètres. La convergence des chaînes de Markov

a été vérifiée grâce au critère de Gelman-Rubin multivarié (Gelman & Rubin 1995), implémenté dans la librairie CODA (Plummer *et al.* 2006) du logiciel R (R Development Core Team 2009).

Résultats obtenus et discussion

- Tout d'abord, le critère de Gelman-Rubin indique une bonne convergence des chaînes entre les 5 chaînes (valeur multivariée de 1.03, valeur par paramètre >1 et < 1.06). En effet, le principe de ce test est de comparer la variance intra chaîne à la variance inter chaînes. On obtient une valeur par paramètre puis une valeur multivariée. Une valeur proche de 1 (aussi bien par paramètre qu'en multivarié) indique une bonne convergence des chaînes.

- Nous avons calculé le facteur de Bayes (fB) pour tester indépendamment les signaux de « fondation » et « d'explosion » sur nos données (Jeffreys 1961; Kass & Raftery 1995). Le facteur de Bayes est un rapport de probabilité postérieure du modèle mesurant le nombre de fois le long de la chaîne où l'évènement A est vérifié / nombre de cas où l'évènement B est vérifié. En considérant l'évènement A comme $N_0 > N_1$ et l'évènement B comme son contraire $N_0 < N_1$ on peut tester le scénario d'expansion. Le principe est le même pour le test du scénario de fondation (comparaison de N_2 avec N_1). On prend ensuite en compte une règle d'évidence pour interpréter les valeurs obtenues. Nous avons obtenu une valeur de fB de 2.2 pour le test d'expansion et de 3.6 pour le test de fondation indiquant respectivement un support faible et modéré de ces scénarios selon l'échelle de Jeffreys. En d'autres termes, MSVAR détecte bien un signal de fondation et d'explosion dans nos données, mais ce signal n'est pas fort.

- Les profils d'estimation des paramètres obtenus sont présentés en figure 9 & 10.

→ Tout d'abord, en théorie, seuls les paramètres mis à l'échelle devraient être interprétés (car dans la théorie de la coalescence, la fonction de vraisemblance dépend de ces paramètres mis à l'échelle et non de façon indépendante des paramètres naturels). Les profils pour θ_1 et θ_2 présentent des distributions assez « étroites » et légèrement décalées par rapport au mode de celle des *priors* suggèrent une bonne estimation de ces paramètres. Le temps depuis le contact ramené à la taille actuelle montre une distribution plutôt de type bimodale, avec l'un des deux modes ayant l'air de tirer la distribution vers l'extérieur de celle du *prior*. On retrouve de ce fait cet effet sur l'estimation de θ_0 .

→ Cependant, on observe également souvent les paramètres naturels car ils fournissent des valeurs estimées plus faciles à interpréter. Dans le cas présent, on se rend tout d'abord compte que nos données n'apportent pas trop d'information sur le taux de mutation (car la distribution n'est pas trop étroite par rapport au prior et car la valeur estimée correspond à la valeur moyenne du prior). Ce résultat est classiquement observé avec MSVAR (Girod *et al.* 2011). La taille estimée actuelle (708 individus) est largement inférieure à la taille ancestrale avant l'évènement (19958 individus) ce qui semble confirmer que la population n'ait pas ré-atteint l'équilibre mutation-dérive. Finalement, la date de l'évènement de fondation estimée (331 générations) est du même ordre de grandeur que la

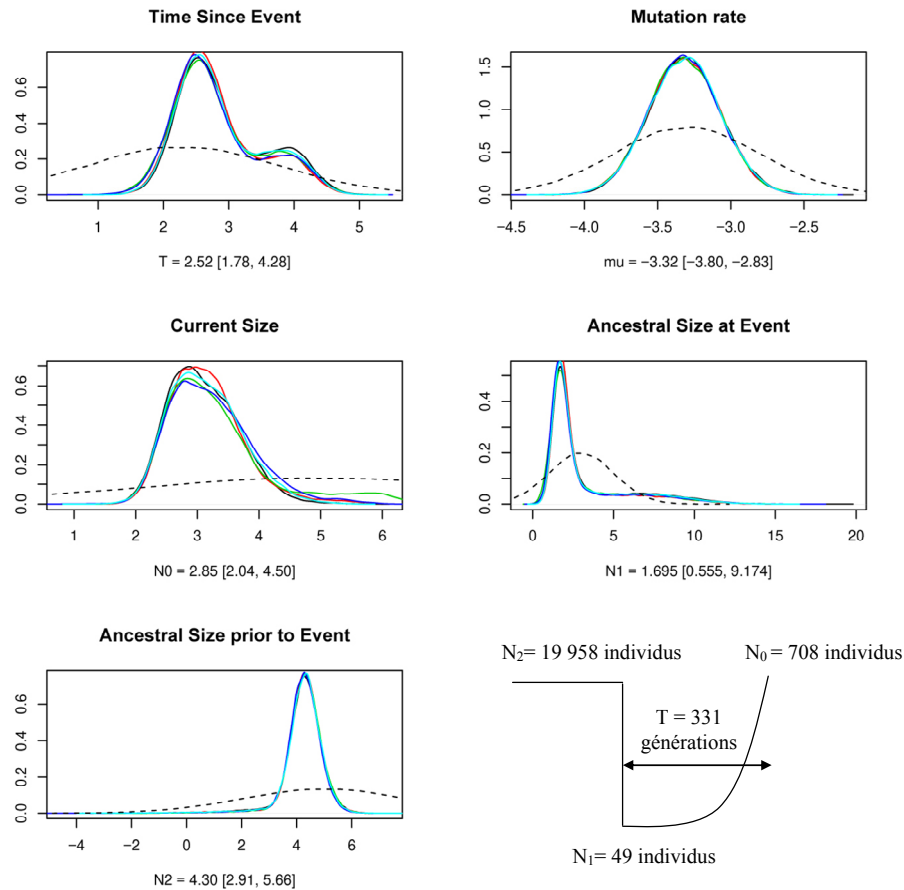


Figure 9. Paramètres naturels estimés par le modèle (les différentes courbes correspondent aux 5 runs et la courbe en pointillés indique la distribution des priors). Les valeurs de paramètres sont en échelles logarithmique. Elles ont été retranscrites en échelle naturelle sur le schéma en bas à gauche de la figure pour clarifier les résultats obtenus. Le taux de mutation estimé est de 5.10^{-4} .

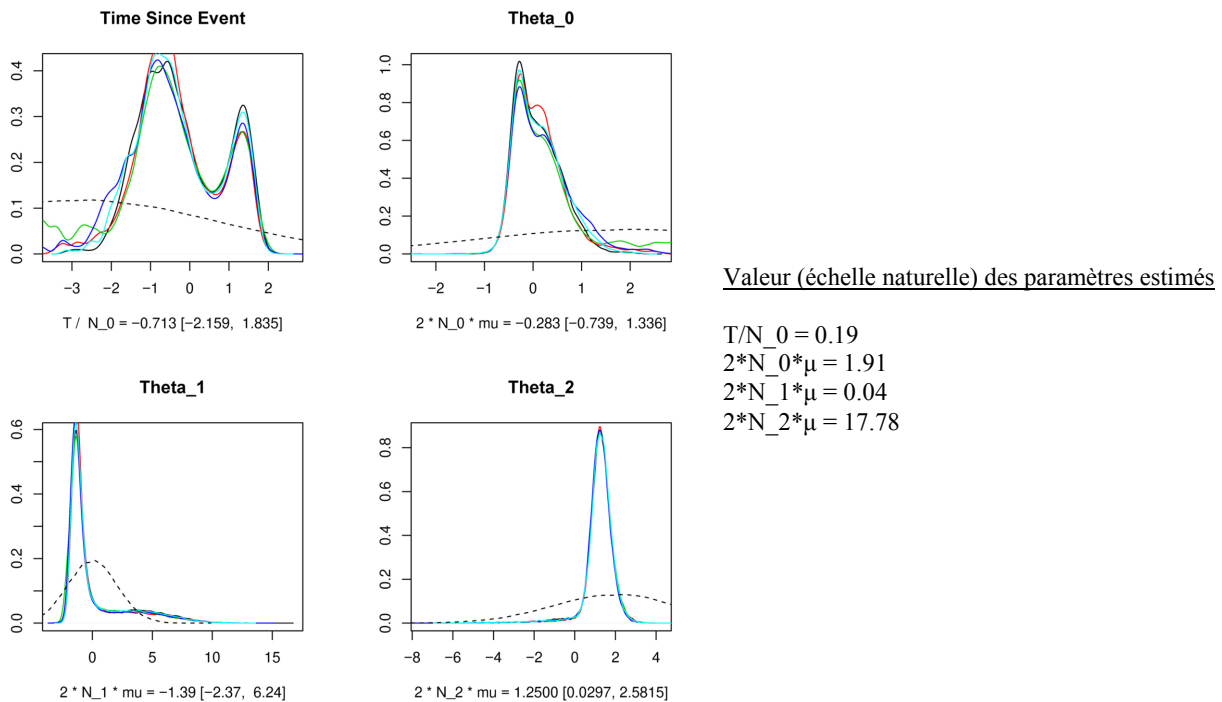


Figure 10. Paramètres mis à l'échelle estimés par le modèle (les différentes courbes correspondent aux 5 chaînes et la courbe en pointillés indique la distribution des priors). Les valeurs de paramètres sont en échelle logarithmique.

valeur historique de 300 générations (considérant la première description de la maladie au Cameroun en 1980 et une génération/mois). La bi-modalité de la distribution observée pour ce même paramètre de temps pourrait s'expliquer par le fait que l'introduction du pathogène au Cameroun a probablement été réalisée à partir d'une population ayant elle-même subi un premier goulet d'étranglement lors de sa première introduction en Afrique. Des analyses plus précises de phylogéographie incluant les différentes populations impliquées sont en cours (Robert et al. soumis) dans le but de spécifiquement tester ce genre d'hypothèses et de retracer les routes mondiales d'expansion de la maladie ainsi que les paramètres démographiques qui y sont associés.

- Les études ayant spécifiquement testé la performance de MSVAR indiquent que certains biais d'estimation peuvent exister. Par exemple, Girod et al. (2011) montrent que la version classique de MSVAR estime assez mal les paramètres actuels dans les situations en « expansion ». En effet, les populations en expansion présentent des généalogies avec des TMRCA (Time to Most Recent Common Ancestor) courts, et donc avec peu de mutations. Dans le cas particulier de la méthode fondation-explosion, Claire Calmet (2002) montre également que certaines déviations au modèle démographique comme un délai avant la reprise démographique après fondation, une croissance non exponentielle ou encore une saturation de l'effectif peu après la fondation peuvent entraîner des baisses de performances du logiciel, aussi bien dans la détection de scénarios que dans l'estimation des paramètres. De telles déviations étant possibles dans le cas de *M. fijiensis*, il est important de rester prudent lors de l'interprétation des résultats.

Conclusion & perspectives

J'ai présenté ici une tentative d'estimation de taille efficace au sein d'une population Camerounaise de *M. fijiensis* par utilisation de l'approche MSVAR. La taille efficace actuelle estimée, bien que potentiellement biaisée, apparaît être relativement faible par rapport au nombre total de reproducteur suspectés en conditions naturelles. De tels écarts sont possibles même à l'équilibre mutation dérive mais ils impliquent de très fortes variances de succès reproducteur. Si on admet qu'il est plus parcimonieux de penser que cet équilibre n'ait pas encore été atteint, on peut s'attendre à ce qu'il faille encore beaucoup de temps pour que la diversité génétique d'origine puisse être récupérée et que l'écart entre la taille efficace et le nombre d'individus reproducteurs ne se réduise. Dans ce contexte, il serait intéressant d'arriver à estimer de façon directe ou indirecte le nombre d'individus reproducteurs par unité de surface afin de pouvoir le comparer à la taille efficace théorique.

Annexe 2 : Présentation d'une analyse d'assignation génétique spatialisée réalisée sur la zone d'étude globale

Contexte : Je présente ici les résultats d'une analyse réalisée avec le logiciel GENELAND (Guillot et al. 2005) sur l'ensemble des individus échantillonnés au sein de la zone d'étude. Cette analyse globale regroupe les échantillons des deux études indépendantes présentées dans l'article 1 et 2.

Matériel et méthodes : Le jeu de données utilisé pour cette analyse contient un total de 902 individus génotypés à 9 marqueurs microsatellites (marqueurs en commun entre les deux études). L'analyse GENELAND a ici été réalisée en prenant explicitement en compte la localisation spatiale des individus ainsi qu'en considérant le modèle génétique « fréquences alléliques non corrélées ». Tous les autres paramètres du modèle ont été fixés comme présenté dans la partie « matériels et méthodes » de l'article n°1.

Résultats et discussion: La figure 11.a représente la distribution de densité postérieure du modèle (illustrant la fréquence pour laquelle la chaîne de Markov a convergé vers la solution $K=x$ pour x variant de 1 à 10). Ce résultat indique que GENELAND détecte deux entités génétiquement distinctes sur la zone d'étude. La figure 11.b représente la probabilité *a posteriori* d'appartenir à la population n°1. Cette probabilité étant forte pour les couleurs rouges et faibles pour les couleurs blanches. La figure 11.c présente les valeurs de F_{ST} par locus et moyennées sur l'ensemble des locus entre les deux populations détectées. Ces résultats sont cohérents avec les deux analyses indépendantes présentées dans les articles 1 et 2. Cependant, la discontinuité génétique est située un peu plus au sud que dans l'article 1. Cela est probablement dû au plus grand nombre d'individus présents dans la population du sud (par rapport à l'analyse présentée dans l'article 1). Puisque l'information spatiale est explicitement prise en compte, on peut s'attendre à ce que la zone de rupture puisse être artificiellement déplacée vers le sud. La valeur moyenne du F_{ST} est proche de celle précédemment obtenue et on retrouve dans les valeurs/locus une grande variance caractéristique du scénario historique suspecté.

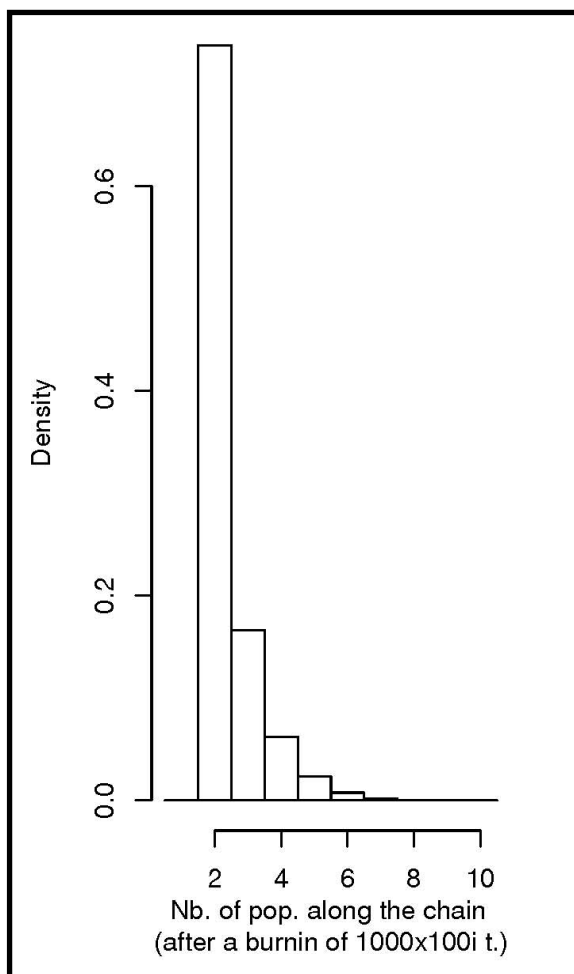


Figure 11.a

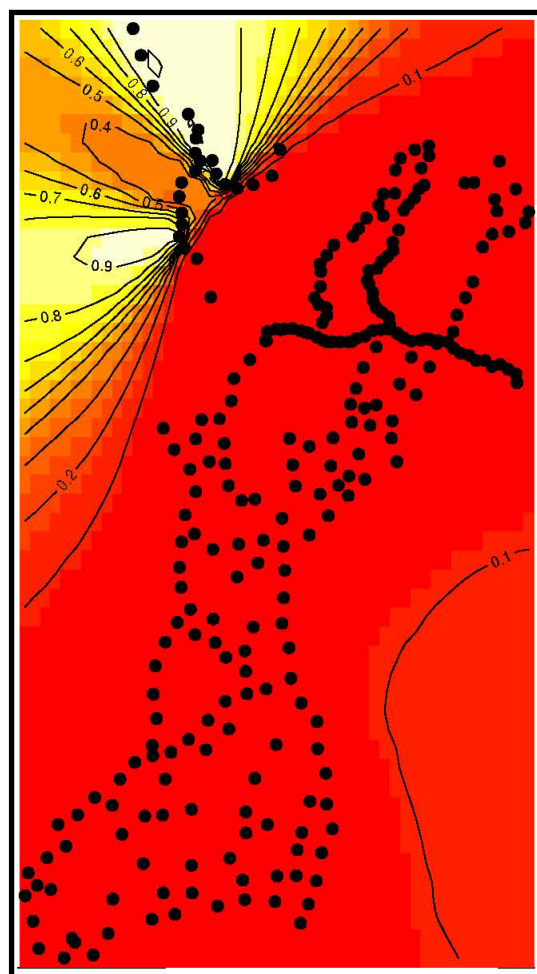


Figure 11.b

Locus	F_{ST}
F62	-0.0018
N194b	0.0492
N137	0.0997
N203	0.1072
Fe09	0.1233
ff09	0.1538
F40	0.1828
FE05B	0.2501
F26b	0.2503
All loci	0.1350

Figure 11.c

Chapitre 2 : Inférence de paramètres à partir des modèles de clines génétiques

Ce chapitre repose sur deux études qui exploitent des situations où il existe des gradients spatiaux de fréquences alléliques (*i.e.*, clines) dans le but d'inférer des paramètres reliés à des forces évolutives. Dans ces deux situations des clines existent pour des raisons différentes mais ils ont en commun la propriété de ne pas être à l'équilibre, c'est-à-dire que les patrons spatiaux de fréquences alléliques évoluent avec le temps.

L'inférence de paramètres par la théorie des clines génétiques est généralement basée sur une procédure incluant deux étapes (Lenormand et al. 1998). La première étape est descriptive. Son objectif est de détecter des patrons de clines génétiques en ajustant des fonctions théoriques aux données de distribution spatiale des fréquences alléliques observées. Cette étape permet également de caractériser les patrons de clines observés grâce à différents indicateurs propres à la fonction théorique sélectionnée (par exemple une pente dans le cas d'une fonction logistique). La seconde étape est théorique. Elle nécessite le développement ou l'utilisation d'un modèle incluant tous les paramètres en jeu (ceux susceptibles de contrôler ou de faire évoluer les valeurs des indicateurs caractérisant les clines, la migration par exemple). Finalement, l'inférence est réalisée en utilisant une procédure statistique visant à déterminer la valeur des paramètres théoriques les plus vraisemblables pour expliquer les patrons de clines observés. Ces deux étapes ont été menées dans la première étude mais pas dans la seconde ou seule la partie descriptive a pu être réalisée pendant cette thèse.

- Article n°3: On the use of temporal variation in neutral genetic clines to estimate gene flow: a case study in a fungal plant pathogen.

Cette première étude se base sur la théorie des clines neutres historiquement développée par Endler (1977) et précédemment présentée en introduction générale (c.f. Figure 3.a). Le concept sous-jacent à cette théorie est que lorsque deux populations rentrent en contact après avoir divergé suite à une phase d'isolement, la rupture spatiale abrupte mise en place dans les patrons spatiaux de fréquences alléliques lors du contact est censée progressivement s'affaiblir par l'effet de la migration s'exerçant entre les deux populations. Ainsi, en faisant l'hypothèse que les populations sont assez grandes pour empêcher la dérive génétique d'agir et qu'il n'y a pas de différence de densité entre les deux populations, Endler (1977) a trouvé une relation entre l'intensité de la migration, le temps depuis le contact initial et la largeur du cline. Nous avons cherché à appliquer cette approche au cas de la discontinuité génétique détectée dans l'article n°1. Cependant, ne connaissant pas le temps depuis le contact initial, nous avons réalisé un autre échantillonnage au niveau de cette zone quelques générations plus tard. Nous avons ensuite adapté le modèle monolocus d'Endler dans le but d'utiliser la vitesse d'aplatissement de clines multilocus (à l'aide de marqueurs microsatellites) entre ces deux échantillonnages pour quantifier l'intensité de la migration. Une attention particulière a été portée à la description et à l'évolution des patrons de clines indépendamment observés aux différents locus.

Cette article sera prochainement soumis (revues potentiellement ciblées : Evolution ou Genetics).

• Article n°4: Spatio temporal distribution and evolution of fungicides resistances alleles at the interface between treated and untreated populations of a plant pathogenic fungus

Cette deuxième étude fait référence à une autre catégorie de patron de clines, celle impliquant l'effet d'une sélection dépendante de l'environnement pour leurs mises en place. Nous avons réalisé deux échantillonnages (en 2009 et 2010) à l'interface entre une plantation industrielle et des plantations paysannes avoisinantes. Ces deux catégories de plantations sont caractérisées par deux types d'environnements ayant été soumis à des pressions sélectives différentes (en termes de traitements fongicides). Dans cette étude nous avons utilisé la théorie des clines pour décrire les patrons spatiaux de fréquences alléliques à deux locus anciennement soumis à sélection. Cependant, faute de temps, nous n'avons pas pu développer un modèle théorique qui aurait éventuellement pu permettre l'estimation relative de l'intensité des différentes forces évolutives en jeu (migration et sélection négative dans le cas présent). En considérant des situations simplifiées nous avons tout de même essayé de déterminer la gamme de valeurs de paramètres réalistes pouvant expliquer les patrons spatio-temporels de fréquence d'allèles résistants aux fongicides observés. Une telle approche nous a permis de mieux comprendre le fonctionnement des populations de *M. fijiensis* au sein de la situation particulière considérée.

Les résultats présentés dans cet article se basent donc sur des analyses descriptives « préliminaires » qui ne m'apparaissent pas publiables en l'état. Comme l'approche est originale dans le contexte de cette thèse, j'ai souhaité inclure les résultats en l'état dans le présent mémoire. Un troisième échantillonnage réalisé en avril 2011 devrait fournir des données supplémentaires qui pourront être intégrées avec les résultats présentés ici dans une nouvelle version du manuscrit. Une estimation de la différence de densités entre des populations traitées et non traitées a été réalisée lors de ce dernier échantillonnage. L'ensemble de ces données pourront, je l'espère, permettre l'inférence de paramètres suite à la réalisation d'un modèle théorique de clines génétiques évoluant sous l'effet conjoint de la sélection négative et de la migration.

On the use of temporal variation in neutral genetic clines to estimate gene flow: a case study in a fungal plant pathogen

A. RIEUX ^{*}, T. LENORMAND [¶], J. CARLIER ^{*}, L. DE LAPEYRE DE BELLAIRE [‡] and V. RAVIGNE ^{*}

^{*} CIRAD, UMR BGPI, Campus international de Baillarguet, TA A-54K, F-34398 [†]

[¶] CEFV-CNRS, UMR 5175, 1919 route de Mende F-34293 [†]

[‡] CIRAD, Persyst, UPR Syst. Banan. Ananas, TA B-26 / PS4, Blvd. de la Lironde, 34398 [†]

[†] Montpellier Cedex 5, France

SUMMARY

Neutral genetic clines theory provides a powerful conceptual framework to estimate the strength of dispersal in natural populations. In the present study, we focused on a geographic area in which two genetically differentiated populations of a banana plant pathogen fungus are distributed along a continuous and homogeneous habitat. We detected genetic clines of allelic frequencies at 8 out of the 15 microsatellites markers studied. We then analysed the temporal evolution of these multilocus genetic clines over 2 years (approx. 15 generations). We showed that all cline centres are coincident and stable in time. In addition, over the studied period, these clines gradually vanished at a common pace. These findings are consistent with the hypothesis that clines evolve under the effect of migration only and that this migration is isotropic along the considered area. Confronting a theoretical model of the temporal dynamics of neutral genetic clines under migration allowed us estimating the intensity of gene flow of the species in this area. Under diffusion approximation and assuming constant migration through time, we found that σ , the so-called variance in axial parent-offspring distance was equal to 1175 metres/generation. Finally, we discuss the robustness of the σ parameter estimated as well as the historical processes that might be at the origin of the observed clinal patterns. This study illustrates that combining approaches to detect sharp breaks in allelic frequencies distribution over space with analyses of genetic clines to better characterise their temporal variation offers a convenient way to gain solid information on gene movements.

INTRODUCTION

Understanding the forces influencing genetic diversity and quantifying the spatial scale of their influence is a central task in evolutionary biology (Dionne et al. 2008). Among these forces, dispersal is of particular interest for at least two reasons. First, dispersal is directly related to the colonisation processes of species in new environments. In the special case of pathogens, the elucidation of individual movements is crucial as it results in disease spread (Milgroom & Peever 2003). Second, dispersal processes determine the level of gene exchange between implanted populations, which condition their evolutionary potential (Slatkin 1987, Lenormand 2002). However, both direct (demographic) and classical indirect (based on genetic) methods of dispersal quantification have been criticised for their relative failures in the estimation of dispersal (Vandewoestijne & Baguette 2004, Broquet & Petit 2009). In such a context, genetic clines theory provides an interesting alternative and powerful framework to estimate the strength of dispersal (Lenormand et al. 1998, Sotka et al. 2006).

Genetic clines are spatial gradients in gene frequencies of a polymorphic genetic marker (Endler 1977). Clines may take several forms, from large areas of overlap to narrow contact zones (Arnold 1992). Generally, such original spatial genetic patterns appear in geographical zones in which genetically differentiated populations meet, have continuous interchange and produce fertile crosses (Barton & Gale 1993). Clines can be at genetic equilibrium, maintained by a balance between the homogenizing effect of dispersal and the diversifying effect of selection (Slatkin 1973, Barton and Hewitt 1985), the latter might be against hybrid genotypes (endogenous selection). However, genetic clines can also appear in a continuous population if individuals in geographically adjacent areas are submitted to different selection pressures (exogenous selection) (Haldane 1948, Nagylaki 1975). In both cases, because of the interaction between migration and selection influence on the spatio-temporal evolution of allele frequencies at selected or linked neutral alleles, genetic clines theory provides a powerful conceptual framework to study both evolutionary forces.

The theory of genetic changes along clines in one and two dimensions has been theoretically and empirically carefully investigated in various species (for review, see Endler 1977, Barton and Hewitt 1985, Barton and Gale 1993, Harrisson 1993). However, most published studies of genetic clines have focussed on tension zones, where endogenous selection plays a major role in allelic frequencies evolution. Besides, most of these studies made use of multilocus selected and/or linked neutral markers to measure clines parameters and linkage disequilibrium with the objective of estimating parameters related to gene flow and selection. In tension zones, unlinked neutral genetic clines have only been used to be compared with clines of selected markers in order to indicate the extent to which the overall genome is under indirect selection (Szymura and Barton 1991, Alexandrino et al. 2005). It is a fact that genetic clines reflecting adaptation to abruptly changing environments (i.e selective pressures) have been much less studied than tension zones. One of the rare examples is the study of genetic clines at selected markers due to local adaptation of the mosquito *Culex pipiens* to the presence of organophosphate insecticides in Southern France (Lenormand et al. 1998, 1999, 2000). Two methods based on genetic cline analysis have been developed from this study in order to estimate parameters related to selection and migration. The

first one is applicable to selected markers clines at genetic equilibrium and requires at least two loci to measure a linkage disequilibrium parameter. In addition, the second one requires at least two temporal samplings and use the rate of change of the shape of a cline to allow parameter estimation in the case of variable migration and selection in time.

The use of genetic cline theory to infer parameters related to dispersal combine the desirable properties of both direct methods, such as measuring dispersal over a specific timescale or inferring dispersal independently from a demographic parameter and indirect methods, such as measuring the effective dispersal of genes and avoiding direct observation of individuals and the achievement of long-term field data. Despite all these valuable advantages, the estimation of dispersal distances through genetic cline theory has seldom been used in comparison to other available methods (Rousset 2001). One of the explanations could be the non existence of user-friendly method or software including all the steps from the characterization of a cline pattern to the estimation of parameters. Another one could be the difficulty to characterize selected or linked neutral markers indispensable for the analysis of genetic clines at equilibrium. Indeed, most of the published studies are making use of several classical indirect methods and unlinked neutral genetic markers (as microsatellites which are the markers of choice in nowadays population genetics studies) to measure dispersal (Broquet & Petit 2009).

In some special situations in which selection does not influence the evolution of allelic frequencies along the cline, unlinked neutral markers may be useful to measure dispersal using genetic cline theory. Such a situation may be found when divergent populations meet after having been separated by some climatic or geological change or following introduction into a new region (i.e., secondary contact). However, contrary to migration-selection cline patterns, such neutral genetic clines cannot be stable through time. Indeed, if two populations meet, and there is no exogenous differential selection between the areas, or selection against hybrid genotypes, each group will spread smoothly into the other's range and the step in allele frequencies produced at the moment of contact will dissolve, leaving only a smooth, gradually falling cline. In this simple scenario, assuming no difference in density between the two groups and population sizes large enough to avoid the effects of genetic drift to become significant, the width of a neutral cline is supposed to depend only on the dispersal rate and the number of generations since contact (Endler, 1977). Despite the apparent simplicity in this neutral approach, no study made use of neutral genetic clines theory to infer parameters related to dispersal yet.

Here, our goal was to measure dispersal abilities of *Mycosphaerella fijiensis*, a wind-dispersed plant pathogen fungus causing black leaf streak disease (BLSD) that is the most destructive leaf disease of banana (Stover & Simmonds 1987). This pathogen has recently spread from South East of Asia to Pacific Islands, Africa and Latin America (Carrier et al. 2000). It has been observed for the first time in Africa in 1973, and in 1980 in Cameroon. In a previous landscape genetics study, Rieux *et al.* (2011) detected the presence of two differentiated populations over space in a South-West Cameroonian banana producing area but they could not associate genetic structure with any landscape/ecological features of the study site. Their results suggested that the sharp genetic discontinuity observed in this area might be explained by the recent history of the pathogen

colonization over the study area (e.g., secondary contact between two expanding populations with a high effective size) rather than contemporary gene flow. Such a situation appears to be interesting to be assessed using genetic clines models with the objective to better characterize *M. fijiensis* dispersal processes, especially since the previous application of classical methods of dispersal inference did not succeed probably because hypothesized large population sizes (Halkett et al. 2010, Rieux et al. 2011). In this species, a better knowledge of these processes is of particular interest since it may help to improve our capacity to predict patterns of disease spread both at local and global scale (Shaw 1995) and to define new efficient management strategy of disease control (Forster & Gilligan 2007).

In this article, we first used a genetic clustering approach to show that the genetic discontinuity previously detected was still present in a second temporal sampling realised 15 generations later. Second we characterized the spatio-temporal evolution of multilocus neutral clines by testing their concordance and coincidence using a method specifically designed to fit clines of molecular genetic data. Finally, we compared the observed multilocus diminishing cline pattern with expectations of a neutral diffusion model built to fit a secondary contact situation. Such an approach allowed us to estimate an average dispersal distance parameter (σ) and its associated confidence interval in *M. fijiensis*.

MATERIALS AND METHODS

Study area and sampling strategy

Mycosphaerella fijiensis samples were collected from untreated plantain leaves infected with BLSD in an agricultural landscape located in South-West Cameroon. The samples were collected in easily accessible areas along the roadside verges. At each sampling site, infected leaf fragments were collected from a few closely spaced banana plants, and GPS coordinates were recorded with a precision of 3 m (WGS84 data format). Two temporal sampling were realized in october 2005 (T₁) and february 2007 (T₂) (see Figure 1 for location and distribution of sampling sites). Considering sexual reproduction, we estimated that 16 pathogen generations might have occurred in the time interval between the two samplings (Jones 2000). Between 3 and 8 individuals were isolated per sampling site using two different methods. In the first sampling, (October 2005), 24 sites were sampled along a 32Km long transect. Mycelium cultures were initiated by single ascospores isolated from necrotic lesions, grown on solid medium (300ml V8, 3g CaCO₃, 20g agar per litre, pH 6) for two weeks at 25°C and dried for 1 day at 55°C before being stored at -20°C until DNA extraction. In the second sampling, (February 2007), 44 sites were sampled along a 45Km long transect and the pathogen isolation step was avoided by collecting young lesions directly from banana leaves (see Robert et al. 2010). The sampled lesions have been individually cut directly after the sampling and dried for 1 day at 55°C before being stored at -20°C until DNA extraction.

Analysis of molecular markers

Total genomic DNA was either extracted from grown mycelium (2005 sampling) or from young lesions (2007 sampling). Both protocols are detailed in Robert et al. (2010). Genotyping was carried out using 3 polymerase chain reaction (PCR) multiplex sets comprising 15 microsatellites loci (Neu et al. 1999, Zapater et al. 2008, Robert et al. 2010). PCR reaction conditions are also detailed in Robert et al. (2010). Among the 15 markers, 9 are tetra-nucleotide, 2 are tri-nucleotide and 4 are di-nucleotide microsatellite loci (see Table 1). Finally, the amplified fragments were separated on a 16-capillary Sequencer (ABI Prism 3130XL, Applied Biosystems). They were sized and analysed for length variation using GeneMapper® Software (Applied Biosystems, Foster City, CA). A total of 151 individuals (2005 sampling) and 254 individuals (2007 sampling) were genotyped using this procedure. DNA samples are stored in the UMR BGPI (Montpellier).

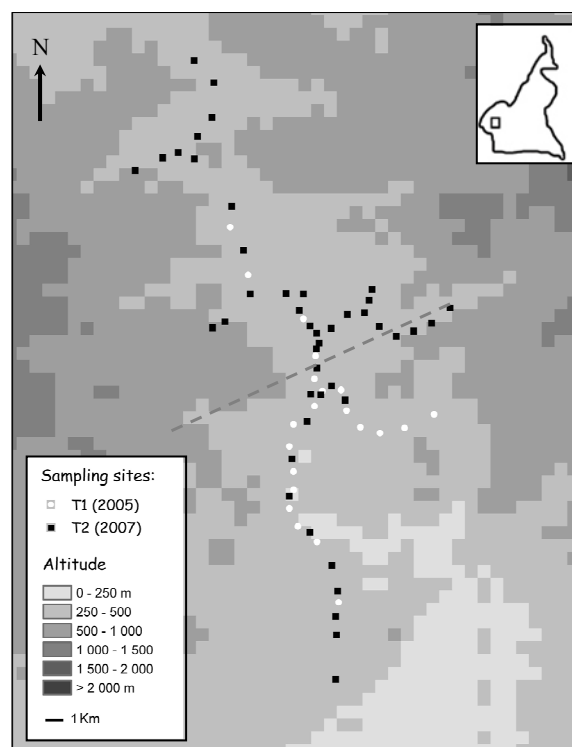


Figure 1. Sample site locations in the south-west Cameroonian North-South transect. The dashed line represents the genetic discontinuity

Delineating genetic discontinuity of the secondary contact zone through clustering approach

Spatial population genetic structure of both data sets was assessed using the clustering algorithm implemented in GENELAND v3.1.4 (Guillot et al., 2005). The basic-unit of this method is individual multilocus genotypes. Subdivision of the data into clusters is done by minimizing Hardy-Weinberg disequilibrium and gametic phase disequilibrium between loci within groups. Geneland can explicitly take spatial localisation information for genotyped individuals into account. GENELAND analyses were conducted using the independent model of allele frequencies and long

MCMC runs to ensure chain convergence (Guillot *et al.* 2005, 2009). We performed five independent runs with 1 000 000 iterations (thinning = 100) allowing K to vary from 1 to 10. The maximum rate of the Poisson process was fixed at 500, the maximum number of nuclei in the Poisson-Voronoi tessellation at 300, and the potential error for spatial coordinates at 2m. K was inferred from the modal value of the run highest likelihood. The run with the highest PPD (posterior probability of the data) was post-processed (with a burn-in of 5×10^4 iterations) in order to obtain posterior probabilities of population membership for each individual and each pixel of the spatial domain.

Estimates of global basic statistics and genetic differentiation between clusters

Gene diversity was estimated using the expected heterozygosity H_E (Nei 1978) and the number of alleles (N_A). These two indices were calculated on the entire data game for both temporal sampling using FSTAT_{2.9.3} (Goudet 1995). We used the non parametric Wilcoxon–Mann–Whitney test to look for significant differences in heterozygosity between the two temporal samplings. Linkage disequilibrium between all pairs of loci on both entire data games was tested using Fisher’s exact tests implemented in GENEPOP (Rousset 2008) run on the web. We used the false discovery rate (FDR) procedure implemented in the R package QVALUE (Storey 2002) to control for multiple testing. The level of genetic differentiation at microsatellite loci among clusters was estimated as F_{ST} (Weir & Cockerham 1984) using GENEPOP. We used the exact likelihood ratio (G) test implemented in this software to test for significant values of F_{ST} between inferred clusters.

Cline fitting tool

We adapted and used the Cfit program version 7 (available at <http://www.cefe.cnrs.fr/ecogev/siteGB/CFitpage.htm>) (Gay et al. 2008) to fit genetic clines using allelic count data. In this approach, loci are considered to have only 2 alleles, let say a_1 and a_2 and the frequency of a_1 can be fitted to a decreasing function (the right allele at each locus should be chosen accordingly) of geographical distance. Several clines can be fitted simultaneously to compare their slopes (concordance) and position (coincidence) and to construct complex hypotheses by constraining some of the parameters involved. Estimation is performed by maximum likelihood using a simulated annealing algorithm that includes Metropolis routines.

Genetic data

In our data game, two alleles were present at 13 of the 15 markers. We reduced the information in the two others markers to two allele systems by assigning each allele to a cluster-specific compound allele (Bierne et al. 2003) according to its coordinates on the first axis of a multiple correspondence analysis (MCA) (using GENETIX, Belkhir et al. 1998). Moreover, because sampling sites were aligned along both the main roadside verges and secondary axes (see figure n°1) we used two-dimensional clines.

Neutral genetic clines fitting strategy

The inference of parameters through genetic cline is generally based on a two steps procedure. The first one is descriptive, relies on the detection and characterization of genetic clines patterns and is usually achieved by fitting theoretical functions to observed data (see steps 1 and 2). The second step is theoretical and requires the use of a model that integrates all the parameters of interests in order to calculate their value under different scenario (see step 3). Maximum likelihood is used to estimate the value of the most likely set of parameters from the theoretical model that best fit the described patterns.

Step 1: cline shape selection

We started by fitting three different shapes named A, B and C (Figure 2) to multilocus clines without exercising any constraints on any parameters in order to determine which one was the best describing the data.

- Shape A is called *stepped cline* (Szymura and Barton 1986) and corresponds to a three-part shape containing a central sigmoid part and two exponential tails. Two parameters are necessary to characterize the central sigmoid (the center c and the slope b) and four parameters are used to characterize the two tails (the distance from the center at which the right and left exponential tails start and their relative slopes, $d1$, $d2$, $t1$ and $t2$ respectively (total of 6 parameters).
- Shape B is called *simple logit* and correspond to the central sigmoid of the stepped cline. It can easily be obtained by using Shape A without fitting tails parameters (Total of 2 parameters, equation B). In such a shape, up and down asymptotic frequencies (i.e., plateaux) are automatically fixed to a value of 1 and 0 respectively.
- Shape C is called *scaled logit* and correspond to a modified sigmoid in which both up and down asymptotic frequencies may take each possible value between 0 and 1 (total of 4 parameters). In this model, h_1 is the lower asymptotic frequency and h_p determines h_2 (the higher asymptotic frequency) as $h_2 = h_1 + h_p (1 - h_1)$. One supplementary parameter should be added to each of these three shapes to fit the angle of the cline resulting in the existence of $k = 3$, 7 and 5 parameters for shape A, B and C respectively. Whereas shape A (and then shape B) were originally implemented in Cfit7, we included shape C in the software, as well as the possibility to fit 2-dimensional clines. Multilocus clines (15L) were independently fitted for allelic frequencies at both temporal samplings (T_1 and T_2). Shapes comparison and choice was done using AIC (i.e., smaller by more than two AIC points) (Raufaste et al. 2005).

Step 2: Locus by locus test of diminishing neutral cline pattern

Considering the best shape, we specifically test each locus for a significant pattern of diminishing neutral cline. First, we compared T_1 sampling fits of each single locus with a model built under the assumption of “slope = 0”. The likelihood of such a null model (corresponding to the worse model possible) was computed locus/locus as:

$$L_0 = Ln(p^{Na_1 + Na_2 + \dots + Na_n} (1-p)^{Na_1 + Na_2 + \dots + Na_n})^{\{1\}}$$

with p the average frequency of allele a over all sites and Na_1 , Na_2 , Na_n the number of alleles a at site 1, 2 and n . Second, for each locus presenting a significant cline pattern at (T_1), we tested for a temporal diminishing cline pattern by comparing the sum of the two independent and no constrained sampling models (T_1 and T_2) with a model in which we constrained the slope to be the same. Likelihood-ratio tests were used to compare the different models and model choice was made at the 5% significance level using a χ^2 test at df difference in freedom degrees between the compared models.

Step 3: Inference of dispersal through the reparametrization of the secondary contact model into a pure diffusion model

In a secondary contact situation, assuming no differences in density between the two groups and high effective population sizes, width of neutral clines (w) are expected to decay after the contact as a function of σ , the variance in axial parent-offspring distance under diffusion approximation and T , the number of generations since contact (Endler, 1977)

$$w = \sqrt{(2\pi)}\sigma\sqrt{T} \quad \{2\}$$

Assuming that the width of a cline $w = \Delta p/s$ with Δp the maximum variation in allelic frequency of allele between the two groups and s the maximum slope of the cline, it is possible to use the temporal variation in cline widths to infer dispersal without any knowledge on the contact time as:

$$\sigma = \frac{\sqrt{w_2^2 - w_1^2}}{\sqrt{2\pi(T_2 - T_1)}} \quad \{3\}$$

with w_1 and w_2 the widths of clines of the first and second sampling respectively and ($T_2 - T_1$) the number of generations in between the two temporal samplings.

We used equation 3 and shape C to build a pure diffusion model in which both temporal sampling could be included. Indeed, we derived the likelihood of such a diffusion model by the addition of the two underlying temporal sampling model likelihoods ($p[T1] + p[T2]$).

The first temporal multilocus cline is described by shape C as:

$$h_1 + h_p(1 - h_1) \frac{e^{b(c-x)}}{1 + e^{b(c-x)}} \quad \{4\}$$

And the second temporal multilocus cline is described by:

$$h_1 + h_p(1 - h_1) \left[1 - \frac{1}{1 + e^{\sqrt{\frac{2(c+\delta-x)}{\frac{4}{b^2} + \frac{1}{2}\pi T\sigma^2}}}} \right] \quad \{5\}$$

Equation 5 has been analytically obtained by calculating the width of a cline at T_2 as a function of the one measured at T_1 and the number T of generations in between T_1 and T_2 . Because the contact zone can move with time, in the case of asymmetric gene flow for instance (Barton & Hewitt 1985) we allowed the centre of the clines to be different between T_1 and T_2 by a shift distance δ . This global model has been implemented in Cfit and was used to fit multilocus clines with the objective of estimating σ . We computed σ support limits within 2 units from the maximum likelihood value. Such a reparametrization allows to fit one width (combination of slope and asymptotic frequencies) by locus and to construct and test other models by constraining some other parameters (c , δ , σ) between locus or temporal samplings to be the same (Table 5). We corrected the likelihood for

overdispersal of genetic data using the variance inflation factor \hat{c} (Lebreton et al. 1992), which was estimated from the residual deviance of the complete model divided by its residual degrees of freedom.

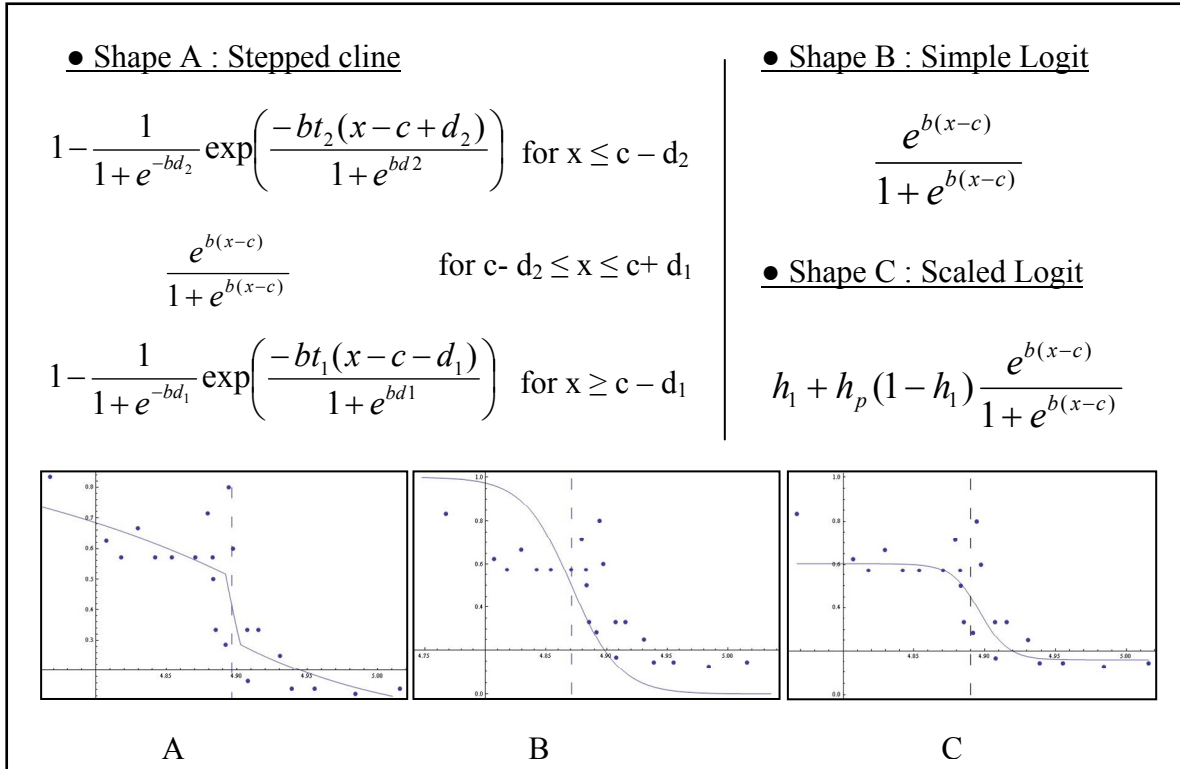


Figure 2. Equations and illustrations of the three shapes considered during the “shape selection fitting step” (see text for details). Illustration is given by plotting the allele frequency of allele 211 at locus 407 (T_1) against geographical distance (the dashed line represents the centre of the cline)

RESULTS

Clustering results

In both 2005 (T_1) and 2007 (T_2) data sets, GENELAND clearly inferred the presence of two clusters (not shown). In T_1 , 86 individuals were assigned to the first cluster and the 56 other individuals were assigned to the second cluster (assignment threshold = 0.80). These two clusters had distinct spatial domains (Figure 3), and the multilocus genetic discontinuity was located at latitude approximately equal to 4.90. Applying the same assignment threshold in T_2 , 158 individuals were assigned to another first cluster and 80 other individuals were assigned to a second one. Finally, five individuals belonging to the same site couldn't be gathered to one or the other cluster as they displayed posterior probability of assignation to both clusters equal to 0.50. The two clusters had also distinct spatial domains (Figure 3). Both T_1 and T_2 maps of population membership posterior probabilities were coherent since the multilocus genetic discontinuity appeared to be located at the same latitude (~ 4.90).

Global genetic diversity and genetic differentiation between inferred clusters

Basics indices of genetic diversity for global population in T_1 and T_2 samplings are summarized in Table 1. When looking at the values of the two temporal sampling, we can detect the presence of new alleles in T_2 in comparison to T_1 . However, the low frequency of these new alleles created by mutation didn't significantly influence the expected heterozygosity between the two temporal samplings ($P=0.815$). No pairs of loci showed significant linkage disequilibrium after FDR control ($Q\text{-value} < 0.05$). Significant genetic differentiation values between the two clusters were observed in both sampling ($F_{ST} = 0.091^{***}$ in T_1 and $F_{ST} = 0.071^{***}$ in T_2). These F_{ST} values varied greatly among microsatellite loci (Table 2), which might be indicative of a secondary contact situation. Indeed, the drastic effects of genetic drift on genetic diversity during founder effects should randomly affect the different loci observed.

locus		T_1 (N=151)		T_2 (N=254)	
		Ar	H_E	Ar	H_E
401	tetra	2	0.179	3	0.193
405	tetra	2	0.464	3	0.397
407	tetra	2	0.499	2	0.466
413	tetra	2	0.495	2	0.502
417	tetra	2	0.484	3	0.445
425	tetra	2	0.254	2	0.371
428	tetra	2	0.482	2	0.502
430	tetra	2	0.055	2	0.094
434	tetra	2	0.478	2	0.485
F40	di	2	0.352	2	0.347
F62	di	2	0.12	3	0.08
FE05B	di	3	0.563	4	0.597
Fe09	di	2	0.455	3	0.475
ff09	tri	2	0.188	2	0.225
N203	tri	3	0.402	3	0.438
All loci		2	0.365	3	0.374

Table 1. Summary of basic information and overall population genetic analyses of both *Mycosphaerella fijiensis* temporal samplings (2005: T_1 & 2007: T_2). N , sample size; H_E , unbiased estimate of gene diversity (Nei 1978); Ar , alleles number are computed among the different microsatellites loci (characterized by a di-, tri- or tetranucleotide repeats)

Locus	F_{ST}	
	T_1	T_2
401	0.016	0.008
405	0.132	-0.001
407	0.220	0.229
413	0.228	0.123
417	0.007	0.060
425	0.268	0.262
428	-0.001	0.100
430	-0.010	-0.010
434	0.116	0.047
F40	0.103	0.092
F62	0.001	-0.008
Fe05B	0.201	0.033
Fe09	0.060	0.109
ff09	0.032	-0.005
N203	-0.005	0.034
All loci	0.091 ***	0.071 ***

Table2. F_{ST} values among the different loci calculated between the two inferred clusters detected using GENELAND for both temporal sampling. (***) = P-value < 0.001 for the Fisher's exact test on the multilocus value).

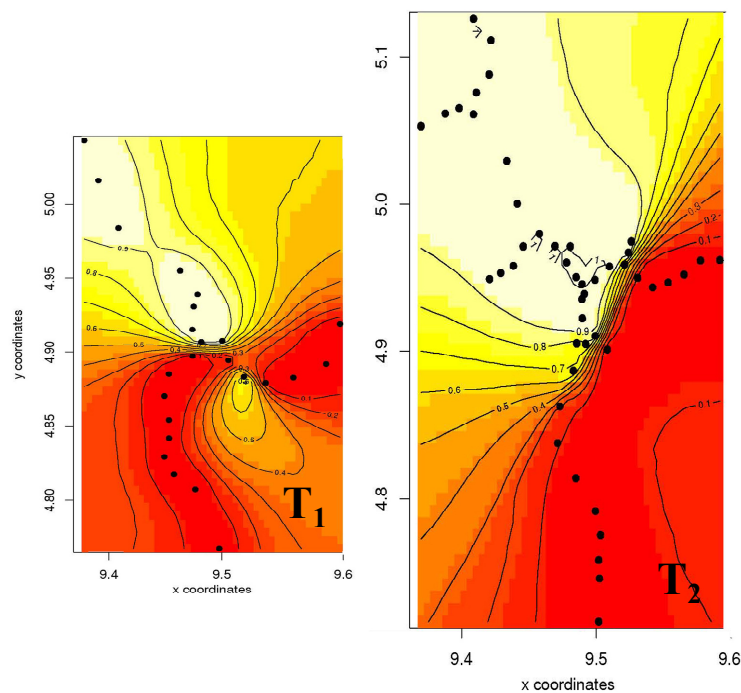


Figure 3. Spatial genetic structure of *Mycosphaerella fijiensis* populations over the study area illustrated by the map of posterior probability of belonging to cluster 1 obtained with GENELAND when analysing T_1 (left) and T_2 (right). Probability value is represented by a white-to-red graded scale; black dots represent sampled sites and y-axis correspond to latitude

Neutral genetic clines fitting

Cline shape selection

Scaled Logit shape performed largely better than the two others (Table 3). This result indicates that for most loci, allelic frequencies are not fixed at the two ends of the geographical transect (e.g. Figure 4). Thus there is a significant gain in likelihood for the scaled logit because both tails at each locus can take any height value.

Locus by locus test of diminishing neutral cline pattern

Approximately half of the loci (8/15) presented a cline pattern (slope significantly different from 0) (Table 4). This result is coherent with the one obtained with F_{ST} values between clusters and is also indicative of a secondary contact. Such a locus/locus fitting approach allowed us to characterise the widths of the several clines obtained at T_1 , widths that appeared to be narrow and locus specific (between 3 and 39 km) since each locus was characterized by its own slope and asymptotic frequencies (Table 4). All the 8 loci presented a diminishing cline pattern. Indeed, the width of the fitted second temporal sampling (T_2) cline was always significantly larger than the one obtained from the T_1 (results not shown, but see Figure 4).

	Shape		
	A	B	C
LL T_1	-1009.0630	-1035.3525	-993.1013
LL T_2	-1651.1543	-1716.7665	-1629.2642
LL Total	-2660.2174	-2752.1191	-2622.3655
Total k	210	90	150
AIC	5740.4347	5684.2381	5544.7311

Table 3. Cline shapes selection result. Multilocus clines (15L) were independently fitted for each temporal samplings. k for number of parameters. Log likelihood and corresponding AIC are computed. Best value is highlighted in bold.

	407	425	405	Fe05b	fe09	413	434	f40	401	430	f62	417	428	N203	ff09
LL T_1 Shape C (K=5) T_1	-84.42	-46.74	-87.17	-83.71	-74.65	-85.89	-83.46	-82.19	-38.45	-16.41	-16.62	-88.39	-91.55	-75.26	-36.14
Angle (radian)	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87
Center	1.86	1.85	1.87	1.87	1.87	1.85	1.84	1.86	1.60	1.72	1.82	1.79	1.78	1.61	1.85
b	68.33	140.19	11.20	25.73	17.25	35.72	38.97	22.92	0.01	0.00	0.02	0.00	0.01	0.00	0.00
Slope	8.79	12.62	2.67	3.77	4.16	5.24	6.29	3.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
h1	0.13	0.62	0.00	0.08	0.02	0.16	0.17	0.00	0.67	0.99	0.96	0.45	0.46	0.12	0.85
hp	0.59	0.94	0.95	0.64	0.98	0.69	0.78	0.70	0.01	1.00	1.00	0.05	0.00	0.10	0.04
h2	0.64	0.98	0.95	0.67	0.98	0.74	0.82	0.70	0.67	1.00	1.00	0.47	0.46	0.21	0.85
Width (Km)	6.44	3.14	39.30	17.10	25.51	12.32	11.29	19.20	x	x	x	x	x	x	x
LL $_0$ Worse model (k=1)	-95.54	-59.02	-93.36	-92.43	-84.69	-94.48	-96.29	-91.61	-38.46	-18.13	-18.13	-88.95	-93.19	-76.76	-37.14
Dev : -2LL(LL $_0$ -LL T_1)	22.24	24.55	12.40	17.44	20.07	17.17	25.65	18.84	0.02	3.46	3.04	1.11	3.29	3.01	2.00

Table 4. Locus/locus cline pattern tests results. Each locus at T_1 was fitted using the scaled logit function. Log likelihood and estimated parameters are computed. For each locus, the non constrained model was compared with the worse model possible to test for a significantly different from 0 value slope (see text for details). Significant values (5% level) are highlighted in bold (Dev for deviance). Finally, the width was calculated for each loci that presented a cline as $w = \Delta p / \text{slope}$ (values are indicated in km). Note that the real slope is not b but $1/4 b (-1 + h1) hp$.

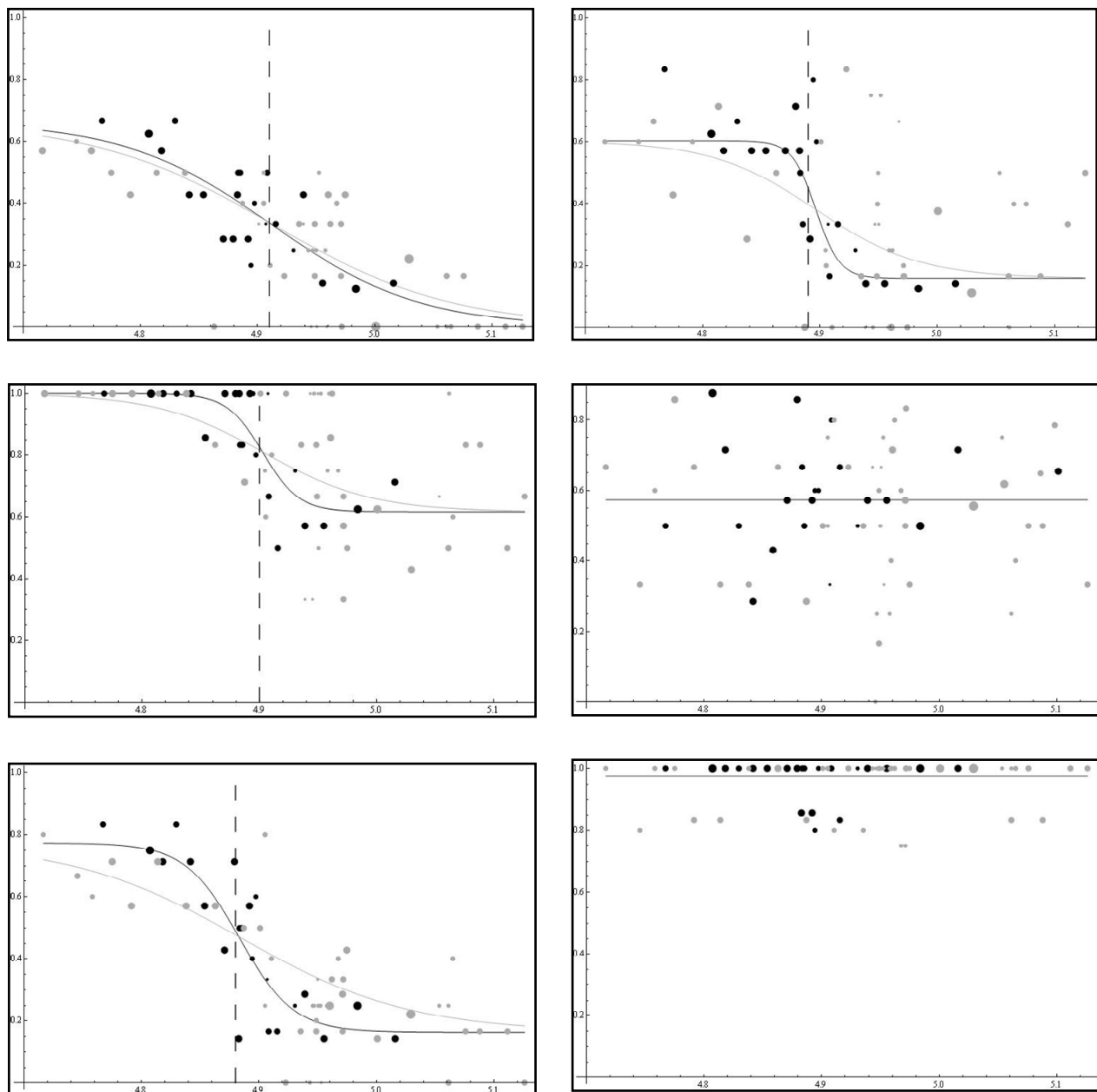


Figure 4. Neutral genetic clines representation at 6 microsatellites (from top to bottom and left to right: 405, 407, 425, 401, 434 and 430). Fits were realised using the scaled logit shape. Allelic frequency is plotted against geographical distance along the cline. Black colours for T_1 and grey colors for T_2 . The size of the points is proportional to the number of individuals sampled and the dashed line represents the centre of the cline.

Inference of dispersal through the reparametrization of the secondary contact model into a pure diffusion model

Two series of multilocus cline adjustments were realized considering the global (T_1 and T_2 are simultaneously fitted) pure diffusion model: the first one on the total data game (15 loci) and the second one on the reduced data game (the 8 loci presenting a diminishing cline pattern). The total data game was used to estimate the strength of σ because all loci should be informative on that point, even those that do not present any cline pattern at T_1 (slope = 0) because they are supposed to present such a null slope at T_2 , which is considered into the theoretical model (i.e., included in the log likelihood) and tested on the data. The reduced data game was used to test for coincidence (common centres between loci) and cline movement through time (through the shift distance δ) because such an assessment could be done on loci presenting a cline pattern only. We first showed that the angle inferred from the 2-D coordinates was not significantly different between the two temporal samplings (Table 5). The inferred value of 1.87 radian (approximately 107°) is consistent with the sampling sites geographical disposition suggesting that dispersal is probably mainly occurring along the North-South sampled transect, transect along which the distribution of the banana plant is the most continuous. Second, we validated the scenario of free diffusion because the model with one sigma (i.e., same σ for each loci) was better than a model in which σ was not constrained to be the same among loci. Finally, the value of σ inferred was 1175 metres/generation and its confidence interval was [636m-1660m] (Table 5 & Figure 5). Analysis realized on the reduced 8L data game provided coherent results on the estimation of σ . The best model was obtained with scenario G that corresponds to a model in which centres are the same between loci at T_1 and between T_1 and T_2 (Table 5).

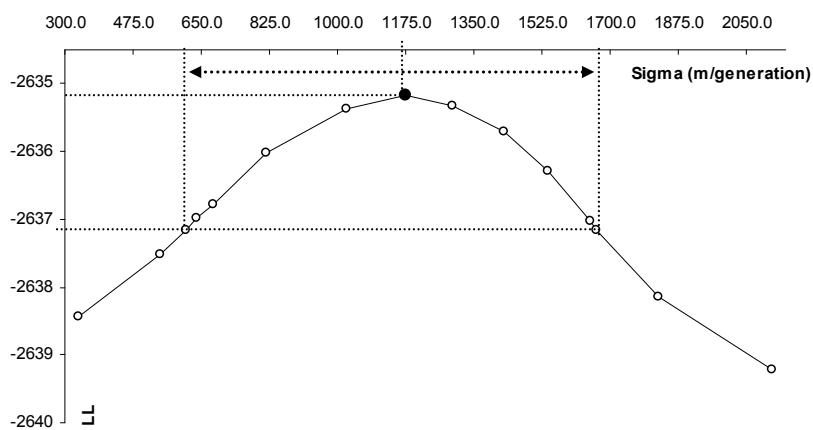


Figure 5. Likelihood profile for σ (m). Dashed line indicates the contour of the 95% confidence region around the maximized likelihood (black point) obtained using model 15L-D.

a)

Model	Parameters constrained
A	$\theta T_1 \neq \theta T_2 / c, b, h_1, h_p, \delta$ and $\sigma \neq$ between locus
B	$\theta T_1 = \theta T_2$ / c, b, h_1, h_p, δ and $\sigma \neq$ between locus
C	$\theta T_1 = \theta T_2$ / c, b, h_1, h_p , and $\delta \neq$ between locus / $\sigma = 0$
D	$\theta T_1 = \theta T_2$ / c, b, h_1, h_p , and $\delta \neq$ between locus / same σ for each locus
E	$\theta T_1 = \theta T_2$ / b, h_1, h_p , and $\delta \neq$ between locus / same σ and c for each locus
F	$\theta T_1 = \theta T_2$ / b, h_1 and $h_p \neq$ between locus / same σ, c & δ for each locus
G	$\theta T_1 = \theta T_2$ / b, h_1 and $h_p \neq$ between locus / same σ and c for each locus / $\delta = 0$
H	$\theta T_1 = \theta T_2$ / c, b, h_1 and $h_p \neq$ between locus / same σ and δ for each locus
I	$\theta T_1 = \theta T_2$ / c, b, h_1 and $h_p \neq$ between locus / same σ for each locus / $\delta=0$

b)

Model	LL	k	AIC	Q_{AIC}	Constrained parameters estimation
15L-A	-2627.2853	92	5438.57	3595.64	
15L-B	-2627.7393	91	5437.48	3594.23	$\theta = 1.8776$
15L-C	-2638.9320	76	5429.86	3578.76	
15L-D	-2635.1655	77	5424.33	3575.87	$\sigma = 1175.69$ m/generation
8L-B	-1606.9448	49	3311.89	2537.94	$\theta = 1.8703$
8L-C	-1611.1645	41	3304.33	2528.34	
8L-D	-1608.8543	42	3301.71	2526.84	$\sigma = 1328.73$ m/generation
8L-E	-1628.4593	35	3326.92	2542.60	$c=1.865$
8L-F	-1616.8349	28	3289.67	2510.95	$\delta = 304.28m$
8L-G	-1617.3972	27	3288.79	2509.81	$\sigma = 1364.75$ m/generation ; $c= 1.8626$
8L-H	-1631.8355	35	3333.67	2547.73	$\sigma = 1172.62$ m/generation ; $\delta= 270.51$
8L-I	-1615.2937	34	3298.59	2520.61	$\sigma = 1364.75$ m/generation ; $c= 1.8626$

Table 5.

a) Identification and construction of the different models tested. Parameters constrained and type of constraint operated are highlighted in bold. (σ for the migration/generation parameter, θ for the 2D angle of the fit, b is related to the maximum slope, h_1 and h_p indicates up & down asymptotic frequencies, c for center and δ for the shift distance between centre at T_1 and T_2).

b) Multilocus (15L or 8L) likelihood search for the different models designed. The best model is highlighted in bold.

DISCUSSION

In a previous study, Rieux *et al.* (2011) localized a sharp discontinuity in the genetic structure of a fungal banana plant pathogen over a wide production area in Southwest Cameroon. Since no correlation between genetic structure and any landscape feature or ecological variable was found, the detected pattern was hypothesized to be due to the recent history of *Mycosphaerella fijiensis* expansion in the area rather than to a current impediment of gene flow (due to e.g., a barrier to dispersal). Specifically, the genetic discontinuity was assumed to stem from either a secondary contact between two expanding populations or a “gene surfing” event. Here, we made use of this interesting situation and of the theory of genetic clines to analyse the temporal evolution of allelic frequencies in this area with the objective to estimate the strength of dispersal. To our knowledge, this is the first study in which this theory has been successfully applied following such an objective. We found only one study in which cline patterns observed at unlinked neutral loci revealed a secondary contact event (Turgeon and Bernatchez 2001). However, the absence of any measured cline data through time impeded the possibility of inferring any parameters from these data. Most other previously published studies focused on the analysis of situations in which selection (either endogenous or exogenous) interacts with migration to influence the spatiotemporal evolution of allelic frequencies along the cline (see Barton and Hewitt 1985, Lenormand *et al.* 1999 for examples on both selection types).

Validity of the neutral assumption

It is known from both cline theory and empirical data that linkage disequilibrium can build up relatively easily in clines maintained by selection (Barton 1983). Studying microsatellite data (*i.e.*, “neutral” markers) therefore does not constitute a guarantee that the observed clines are not affected by selection and evolve under the action of migration only. In the present study, we had *a priori* reasons to argue that the observed genetic clines are purely neutral. First, the sampled area that is relatively small (45 x 20 Km) in comparison to other studied zones is characterized by an apparent absence of any obvious environmental or exogenous selective gradient. The sampled area is composed of a shallow valley where plantain banana trees are extensively cultivated for local consumption in small plots or backyard gardens without any use of fungicides. The distribution of host trees is spatially homogeneous and continuous along the valley. Second, contrary to hybrid zones in which distinct diploid races or sub species meet, all sampled individuals are haploid and belong to the same species, which should limit the occurrence of endogenous selection against “hybrids”.

In addition to these *a priori* indications, the neutral nature of the observed clines was also supported by the results. First the 8 clines identified in 2005 were all shown to significantly smooth out two years (~ 16 generations) later, suggesting that the genetic discontinuity is not at migration-selection equilibrium. No heterogeneity in the rate of increase in cline width was detected among the 8 loci. This is consistent with the neutral assumption. Indeed it has been shown that if physically linked to a selected marker, neutral loci (e.g., microsatellites) should see their cline width increase at a rate

that depends on the strength of selection against introgression at the selected loci with which they are associated (Barton & Gale 1993). Selection is therefore expected to result in contrasted rates of cline width increase at different microsatellites depending on their genetic distance with the loci under selection, which was not observed. Finally, the 7 loci that showed no clinal pattern at the first sampling behaved consistently with the neutral assumption to the extent that no discordant geographic pattern has appeared 16 generations later. As a conclusion, although there is no way to completely exclude a role for selection in the existence of the observed clines, all above observations are congruent with a pure-migration scenario.

On the inference of dispersal strength

From the observation that genetic clines exist at 8 microsatellite loci and fade out at a common rate, it was possible to estimate the strength of dispersal under diffusion approximation (Endler 1977). Such a temporal comparison of clines has allowed previous parameters inference in the case of clines observed at markers under exogenous selection (Lenormand et al. 1999). Theoretical neutral clines simulations have only considered the case in which encountering populations presents fixed allelic frequencies (i.e., $p = 0$ and $q = 1$) in both sides of the contact point (Endler 1977, Durrett et al. 2000). However, here we showed that in natural situations, allelic frequencies are not necessarily fixed and might take any value depending on the locus/locus genetic diversity existing at the event time (contact or surfing). As a result, the theoretical function considered in this study (i.e the scaled logit) appeared to be adapted to be fitted on such genetic data. This approach allowed us to provide the first estimate of a parameter related to the strength of *M. fijiensis* dispersal from genetic data ($\sigma = 1175\text{m} / \text{generation}$ [636m-1660m]). This inferred value of σ may be seen as a synthetic diffusion parameter in which gene flow is averaged over 16 generations. No movement of cline centres was detected, which is consistent with an isotropic dispersal of *M. fijiensis* over the study area and an absence of spatial trends in population density (Nagylaki 1975, Barton & Hewitt 1985). These estimates are supposed to be valid for the area, the period and the scale considered in the present study only. However, despite the inexistence of previous estimates of dispersal in *M. fijiensis* and related species, the inferred value appears to be consistent with the biology of the fungus. Indeed, this pathogen exhibits two natural modes of dispersal. The first relies on asexual spores called conidia that are suspected to disperse at very short distances (Gauhl 1994) whereas the second is ensured by light sexual spores, namely ascospores that are actively projected in the air (Gauhl 1994, Yafetto et al. 2008) and wind dispersed (Burt 2003). As a result, airborne fungus has long been suspected to display important dispersal abilities (Brown 1997). For instance, Burt et al. (1998) and Amil et al. (2007) argued that *M. fijiensis* ascospores might be transported over several kilometres. Combined with important effective size, such important dispersal abilities are consistent with the recent undetection of any IBD pattern along a 33Km long transect (Rieux et al. 2011) and with the present estimation of σ . Nevertheless, it is important to keep in mind that the neutral cline approach presented here as well as most of genetic clines previously published studies relies on the underlying assumption that dispersal is a pure diffusion process (but see Barton 2000 or Saccheri et al. 2008 for exceptions). Such diffusion roughly corresponds with a Gaussian dispersal kernel. However, for wind-dispersed spores, dispersal has been shown to be leptokurtic, i.e., there is an

excess of small and long distance dispersal events in comparison to expected values under a Gaussian distribution (Lacey 1996, Wingen et al. 2007, Sackett and Mundt 2007). In such a context, it is not clear whether leptokurtic dispersal distribution could involve deviations in the estimation of dispersal parameters through genetic clines approaches since this point has been little discussed in the literature. Indeed, we found only two numerical examples investigating the effects of different long distance dispersal amounts on patterns of clines of selected alleles (Lenormand 1998, Rousset 2001). In both cases, it has been shown that large values of kurtosis may significantly increase the maximum slope at the centre of the cline in comparison to a Gaussian dispersal distribution, which could potentially result in the over-estimation of the synthetic σ parameter. Further theoretical work on this point is needed in the specific case of neutral genetic clines.

Back on the origin of clinal genetic patterns

A sharp genetic discontinuity in neutral allele frequencies such as the one observed in Cameroon (Halkett et al. 2010, Rieux et al. 2011) may naturally be explained by a secondary contact between two differentiated populations. In principle, in the case of a secondary contact it is possible to estimate the product $\sigma\sqrt{T}$ where σ is the diffusion coefficient and T the time since contact in generations (see Eq.2, Endler 1977). Hence from our data, where the two temporal samples provide a way to infer σ , we ideally would have been in a position to jointly infer both σ and T . This would only have required writing the likelihood of the shape of the 2005 clines as a function of both parameters. However such analysis was not possible in the present case for different reasons.

The first is that we were not able to validate the scenario of a secondary contact. Indeed, the population structure of the study area is compatible with two scenarios (Halkett et al. 2010; Rieux et al. 2011). The first combines the effect of independent founder effects caused by the movement of a few individuals by long-distance spore dispersal (LDD or movement of infected material) as already suggested from analysis at the continental scale (Rivas et al. 2004), followed by natural gradual diffusion (GD). The genetic structure observed would then effectively be attributable to secondary contact between two expanding patches. The second scenario is based on recent theoretical studies, that have shown that low-frequency alleles can sometimes ‘surf the wave’ of the advance of a population range expansion, reaching high frequencies and spreading over large areas (Edmonds et al. 2004; Bialozyt et al. 2006; Excoffier & Ray 2008). This process, called ‘gene surfing’, can be seen as a combination of drift and spatial expansion. It has already been observed empirically in cultured bacterial populations (Hallatschek & Nelson 2008), but not yet in natural populations.

Distinguishing between these two scenarios is not possible based on historical data only as these are pretty scarce. The earliest report of *M. fijiensis* in Africa was in 1973 in Zambia (Jones 2000). It arrived in Cameroon in 1980 from the adjacent Gabon and was restricted for a while to a Southeastern area near the coast (Kribi region; Foure & Lescot 1988) before being reported in 1983 south (~50 Km) to the study area. In the study area, we could find no information on the spread of BLSD in Cameroon. It is only known that the disease was described northern to the study site in 1986 in Nigeria.

Using genetic data to distinguish between the two scenarios is also tricky. Indeed, whether clear predictions are available concerning clines issued from secondary contacts (Endler 1977) the case of gene surfing has never been considered yet. In particular, at contact time under the former scenario cline centres should not differ among loci and cline slopes are expected to be infinite at all loci exhibiting differentiation between the two populations. In our data, cline centres were found to be coincident, and cline slopes were still steep at some loci in 2005, which is consistent with a secondary contact (Barton & Hewitt 1985, Durrett et al. 2000) without excluding gene surfing.

In contrast, under pure migration, clines formed subsequently to a secondary contact should have the same width after a given number of generations, which is not what we observed. Cline widths in the first sampling ranged from 3 to 39 Km and significantly differed. Such variation may be consistent with gene surfing but we argue here that it could also be compatible with secondary contact in the specific historical case of *M. fijiensis* in Cameroon. Indeed, the species is self-incompatible (haploid heterothallic) so that the production of ascospores requires important enough population sizes for two sexually compatible lesions to spread and collapse on a same banana leave. It is therefore possible that the initial spatial spread of the species at the scale of the study area occurred mainly by short-range diffusive dispersal accompanied by intense genetic drift due to Allee effects over reproduction, which may produce variability in initial cline shape trajectories over time between loci.

In conclusion, whether the clinal genetic variation observed stems from a secondary contact or gene surfing, still has to be elucidated and it is likely that this task requires additional theoretical work. In the present study, variation in initial cline widths suggests a major role of drift during the initial stages of cline formation, which hampers the use of the pure-migration model of cline shape dynamics for inferences based on the first cline shape.

Conclusion

Our results show that combining genetic clustering approaches to detect sharp breaks in allelic frequencies distribution with analyses of genetic clines to better characterize their spatial variation offers a convenient way to gain solid information on gene movements. In such a context, neutral genetic clines theory can prove very useful because it combines the desirable properties of both direct and indirect methods and because predictable frequency patterns are expected and can be specifically tested. Moreover, high effective sizes that are usually considered as a population characteristic limiting the study of dispersal through classical indirect methods (Giraud et al. 2008, Gauffre et al. 2008, Rieux et al. 2011), appears to be a valuable advantage in the application of the neutral genetic cline theory. As a result, we hope that the approach developed in the present paper will be useful in future studies aiming to estimate dispersal parameters, especially in the case of plant pathogens species. Finally, we have estimated high level of gene flow in *M. fijiensis* that may represent an important evolutionary potential at the population level (McDonald & Linde 2002). Such a characteristic should explicitly be taken into account in the future development of new chemical disease control strategies at the landscape scale.

REFERENCES

- Alexandrino J, Baird SJE, Lawson L, *et al.* (2005) Strong selection against hybrids at a hybrid zone in the *Ensatina* ring species complex and its evolutionary implications. *Evolution* **59**, 1334-1347.
- Amil AF, Heaney SP, Stanger C, Shaw MW (2007) Dynamics of QoI sensitivity in *Mycosphaerella fijiensis* in Costa Rica during 2000 to 2003. *Phytopathology* **97**, 1451-1457.
- Arnold ML (1992) Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* **23**, 237-261.
- Barton N (1983) Multilocus clines. *Evolution* **37**, 454-471.
- Barton N, Hewitt G (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.
- Barton NH (2000) Estimating multilocus linkage disequilibria. *Heredity* **84**, 373-389.
- Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process*. (ed. eds RGH). Oxford University Press, Oxford.
- Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (1998) GENETIX, logiciel sous Windows TM pour la génétique des populations. *Laboratoire Genome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France*.
- Bialozyt R, Ziegenhagen B, Petit RJ (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology* **19**, 12-20.
- Bierne N, Borsa P, Daguin C, *et al.* (2003) Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology* **12**, 447-461.
- Broquet T, Petit EJ (2009) Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology Evolution and Systematics* **40**, 193-216.
- Brown JKM (1997) *Plant pathogens and plant diseases* Australasian Plant Pathology Society Inc.
- Burt PJA (2003) Airborne dispersal of *Mycosphaerella fijiensis*. In: *Mycosphaerella Leaf Spot Diseases of Bananas: Present Status and Outlook. Proceedings of the Workshop on Mycosphaerella Leaf Spot Diseases, San José, Costa Rica*.
- Burt PJA, Rutter J, Ramirez F (1998) Airborne spore loads and mesoscale dispersal of the fungal pathogens causing Sigatoka diseases in banana and plantain. *Aerobiologia* **14**, 209-214.
- Carlier J (2000) Population genetic structure and dispersal of *Mycosphaerella fijiensis*. *Info musa* **13**, 17-20.
- Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology* **17**, 2382-2396.
- Durrett R, Buttel L, Harrison R (2000) Spatial models for hybrid zones. *Heredity* **84**, 9-19.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 975-979.
- Endler JA (1977) Geographic variation, speciation, and clines. *Monographs in Population Biology* **10**, 1-246.
- Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution* **23**, 347-351.

- Forster GA, Gilligan CA (2007) Optimizing the control of disease infestations at the landscape scale. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 4984-4989.
- Fouré E (1988) Stratégies de lutte contre la cercosporiose noire des bananiers et des plantains provoquée par *Mycosphaerella fijiensis* Morelet. L'avertissement biologique au Cameroun. Evaluation des possibilités d'amélioration. *Fruits* **43**, 269-274.
- Gauffre B, Estoup A, Bretagnolle V, Cosson JF (2008) Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology* **17**, 4619-4629.
- Gauhl F (1994) Epidemiology and ecology of black sigatoka (PhD Thesis), p. 120. INIBAP, Montpellier.
- Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models *Evolution* **62**, 2789-2806.
- Giraud T, Enjalbert J, Fournier E, Delmotte F, Dutech C (2008) Population genetics of fungal diseases of plants. *Parasite-Journal De La Societe Francaise De Parasitologie* **15**, 449-454.
- Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.
- Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology* **18**, 4734-4756.
- Haldane JBS (1948) The theory of a cline. *Journal of Genetics* **48**, 277-284.
- Halkett F, Coste D, Rivas Platero GG, *et al.* (2010) Isolation by distance and genetic discontinuities in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Molecular Ecology*.
- Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theoretical Population Biology* **73**, 158-170.
- Harrison RG (1993) *Hybrid Zones and the Evolutionary Process* Oxford University Press.
- Jones DR (2000) Diseases of banana, abaca and enset. CABI Publishing, Oxon.
- Lacey J (1996) Spore dispersal - Its role in ecology and disease: The British contribution to fungal aerobiology. *Mycological Research* **100**, 641-660.
- Lebreton JD, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals - a unified approach with case-studies *Ecological Monographs* **62**, 67-118.
- Lenormand T (1998) *Dynamique de l'adaptation locale: Résistance aux insecticides chez Culex pipiens*, Ecole Nationale Supérieure Agronomique de Montpellier
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**, 183-189.
- Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* **400**, 861-864.
- Lenormand T, Guillemaud T, Bourguet D, Raymond M (1998) Evaluating gene flow using selected markers: A case study. *Genetics* **149**, 1383-1392.
- Lenormand T, Raymond M (1998) Resistance management: the stable zone strategy. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 1985-1990.

- Lenormand T, Raymond M (2000) Analysis of clines with variable selection and variable migration. *American Naturalist* **155**, 70-82.
- McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349-+.
- Milgroom MG, Peever TL (2003) Population biology of plant pathogens - The synthesis of plant disease epidemiology and population genetics. *Plant Disease* **87**, 608-617.
- Nagylaki T (1975) Conditions for the existence of clines. *Genetics*, 595-615.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-590.
- Neu C, Kaemmer D, Kahl G, Fischer D, Weising K (1999) Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. *Molecular Ecology* **8**, 523-525.
- Raufaste N, Orth A, Belkhir K, *et al.* (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society* **84**, 593-616.
- Rieux A, Halkett F, de Bellaire LD, *et al.* (2011) Inferences on pathogenic fungus population structures from microsatellite data: new insights from spatial genetics approaches. *Molecular Ecology* **20**, 1661-1674.
- Rivas GG, Zapater MF, Abadie C, Carlier J (2004) Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis*. *Molecular Ecology* **13**, 471-482.
- Robert S, Rieux A, Argout X, Carlier J, Zapater MF (2010) Optimized genotyping with microsatellite markers in the fungal banana pathogen *Mycosphaerella Fijiensis*. *American Journal of Botany* **97**, E130-E132.
- Rousset F (2001) Genetic approaches to the estimation of dispersal rates. . In: *Dispersal*, pp. 18-28.
- Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- Saccheri IJ, Rousset F, Watts PC, Brakefield PM, Cook LM (2008) Selection and gene flow on a diminishing cline of melanic peppered moths. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16212-16217.
- Sackett KE, Mundt CC (2005) The effects of dispersal gradient and pathogen life cycle components on epidemic velocity in computer simulations. *Phytopathology* **95**, 992-1000.
- Shaw MW (1995) Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance *Proceedings of the Royal Society of London Series B-Biological Sciences* **259**, 243-248.
- Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733-756.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.
- Sotka EE, Palumbi SR (2006) The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology* **87**, 1094-1103.
- Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**, 479-498.
- Stover RH, Simmonds NW (1987) Bananas. Willey & Sons, New York.

- Szymura JM, Barton NH (1991) The genetic-structure of the hybrid zone between the fire-bellied toads *bombina-bombina* and *b-variegata* - comparisons between transects and between loci. *Evolution* **45**, 237-261.
- Turgeon J, Bernatchez L (2001) Clinal variation at microsatellite loci reveals historical secondary intergradation between glacial races of *Coregonus artedii* (Teleostei : Coregoninae). *Evolution* **55**, 2274-2286.
- Vandewoestijne S, Baguette M (2004) Demographic versus genetic dispersal measures. *Population Ecology* **46**, 281-285.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
- Wingen LU, Brown JKM, Shaw MW (2007) The population genetic structure of clonal organisms generated by exponentially bounded and fat-tailed dispersal. *Genetics* **177**, 435-448.
- Yafetto L, Carroll L, Cui YL, *et al.* (2008) The Fastest Flights in Nature: High-Speed Spore Discharge Mechanisms among Fungi. *Plos One* **3**.
- Zapater MF, Duchemin M, Dussart JF, *et al.* (2008) Microsatellite markers for the fungal banana pathogens *Mycosphaerella fijiensis*, *Mycosphaerella musicola* and *Mycosphaerella eumusae*. *Molecular Ecology Resources* **8**, 1121-1125.

Spatio temporal distribution and evolution of fungicide resistance alleles at the interface between treated and untreated populations of a plant pathogenic fungus.

A. RIEUX^{*}, L. DE LAPEYRE DE BELLAIRE[¶], J. CARLIER^{*}, A. MEHL[†] and V. RAVIGNE^{*}

^{*} CIRAD, UMR BGPI, Campus international de Baillarguet, TA A-54K – Montpellier, France

[¶] CIRAD, Persyst, UPR Syst. Banan. Ananas, TA B-26 / PS4 – Montpellier, France

[†] Bayer CropScience AG Alfred-Nobel-Str. 50, Building 6240 D-40789 Monheim, Germany

SUMMARY

The evolution of pesticide resistance provides some of the most striking examples of Darwinian evolution occurring over a human life span. In this context, following the spatio-temporal distribution of fungicides resistance alleles at the interface between treated and untreated areas provide an interesting framework to better understand the relative magnitude of the evolutionary forces in game. In the present study, we have investigated the case of local adaptation of the banana plant pathogen *Mycosphaerella fijiensis* to two families of systemic fungicides at a local scale including i) a commercial farm where these fungicides have been previously used and ii) untreated plots in the vicinity. We showed that approximately 50 generations after the systemic fungicide application interruption, previously treated populations still presented detectable levels of resistance to both fungicides. We used simplified theoretical considerations to evoke that both weak values of fitness penalties and/or weak effective migration rates from untreated to treated populations would have been enough to explain the observed temporal decrease in resistant frequencies inside the treated area. Further work is needed to disentangle the relative effect of both forces.

INTRODUCTION

The evolution of pesticide resistance provides some of the most striking examples of Darwinian evolution occurring over a human life span (Lenormand et al. 1999). Indeed, for natural populations, pesticide application represents a strong exogenous selective pressure that has induced the emergence of resistant strains in many well-described situations (Agrios 1997). Because such a selective pressure often abruptly varies in space, it potentially results in the rapid apparition of spatial patterns of resistance allele frequency (*i.e.*, local adaptation). Typically, the frequency of resistant alleles may be very high in treated areas and more or less gradually fall down to low values when going into untreated areas. Such a geographical pattern of spatial variation of allele frequency is called genetic cline. This definition has first been introduced by Fisher (1937) as he was studying the wave of advance of favourable genes. Later, Slatkin (1973) and Nagylaki (1975) have theoretically investigated the conditions for the existence of genetic clines under exogenous selection. An important result was that genetic clines could be stable in time because of the opposite influences of selection and migration. For clines at pesticide resistance genes, selection favours the resistant allele in the treated area and may penalize the resistant allele in untreated areas when a resistance cost exists, *e.g.*, because of pleiotropic effects that might be associated with the resistant allele (Fisher 1930, Caspari 1952). In this system, gene flow between treated and untreated areas counteracts the diversifying effect of selection by introducing genes into the habitat in which they are not adapted (*i.e.*, migration load) (Lenormand 2002).

The theory of migration-selection clines provides a powerful conceptual framework to estimate the relative magnitude of each force in game. Such an approach is generally based on a two-step procedure. The first one is descriptive and relies on the detection and characterization of a cline pattern at selected markers. The second step is theoretical and requires developing a mathematical model that integrates the interaction between gene flow and selection (as well as other parameters of interest such as densities of populations in space, interactions between alleles, or between loci) in order to calculate allelic frequencies under different scenarios. Finally, a statistical approach is used to estimate parameter values from the theoretical model that best fit the described clines. However, when genetic clines are at genetic equilibrium (*i.e.*, maintained by the interaction between migration and selection), disentangling the relative influence of both forces on a cline is not trivial and in all cases, extra information about the system is required. First, an independent measure of dispersal or selection can be used if available (Endler 1977, Barton & Hewitt 1985). Second, a genotypic parameter such as heterozygote deficiency at one-locus or linkage disequilibria when several loci are involved might be useful to provide information on the intensity of migration alone (Mallet & Barton 1989). As an alternative to clines at migration-selection equilibrium, one may study the temporal variation of clines (Lenormand et al. 2000). In principle, because migration and selection do not affect cline shape in a similar manner it is possible to estimate the two parameters jointly. However, despite its potential for the estimation of parameters at a micro evolutionary scale and for the understanding of pesticide resistance dynamic, genetic clines theory has mainly been studied in the case of tension zones (Barton & Hewitt 1985).

One of the rare applications of such approaches to a gradient of pesticide resistance gene has been previously realised in the case of the mosquito *Culex pipiens* local adaptation to organophosphate insecticides in the Montpellier area, south of France (Lenormand et al. 1998, 1999, 2000). These studies have allowed the development of several methods of genetic cline analysis adapted to the pesticide resistance model specificities (in comparison to tension zones) such as the fact that selective pressures are i) generally strong and acting on a few loci, ii) operated in a “finite” environment and iii) often varying in time. As a result, an accurate characterization of the organophosphate insecticides resistance dynamic between treated and untreated area allowed both the estimation of migration and selection coefficient and the definition of new resistance management strategies (Lenormand & Raymond 1998). However, subsequently to those powerful developments, very few studies attempted to characterize such genetic clines *in natura* in other species, probably because i) identifying specific selected markers is not trivial, ii) the close proximity between treated and untreated area is rare in natural ecosystems and iii) the methods are not really “user-friendly” (e.g., there is no software including all the steps from the characterization of a cline pattern to the estimation of migration or selection parameters).

In the present study, we have investigated the case of local adaptation of the banana plant pathogen *Mycosphaerella fijiensis* to two families of systemic fungicides at a local scale including i) a commercial farm where these fungicides have been used and ii) untreated plantain banana farms in the vicinity. The studied area is supposed to provide an interesting testing ground for evaluating key parameters involved in fungicide resistance. Indeed, from the end of the 80’ to 2004, fungicide resistance alleles have been positively selected by the use of systemic fungicides. The situation changed in 2004 when the emergence of too high levels of resistant strains in the commercial farms resulted in the withdrawal of systemic fungicides and their replacement by unspecific contact fungicides that did not select for resistance anymore (de Lapeyre de Bellaire *et al.* 2010). Under such circumstances it can be expected that genetic clines at systemic fungicides resistance alleles have built up until 2004 (see Figure 1.a). Such cline patterns should be now in the way to vanish under the joint action of migration between treated and untreated areas and a potential counter-selection (Figure 1.b). Here we investigate such a hypothesis by analysing the temporal evolution of resistant alleles frequencies for both Strobilurins and Benzimidazoles fungicides families with the objective to estimate the intensity of the underlying forces in game. Mainly by lack of time we did not fully used the cline patterns obtained to estimate migration and fitness costs, as this would have required to build a model of cline evolution under the action of a fitness cost, asymmetric migration and unbalanced population sizes. Instead, as a preliminary study, we furnish parameter estimates compatible with the evolution of allelic frequencies in treated and untreated areas in simplified situations.

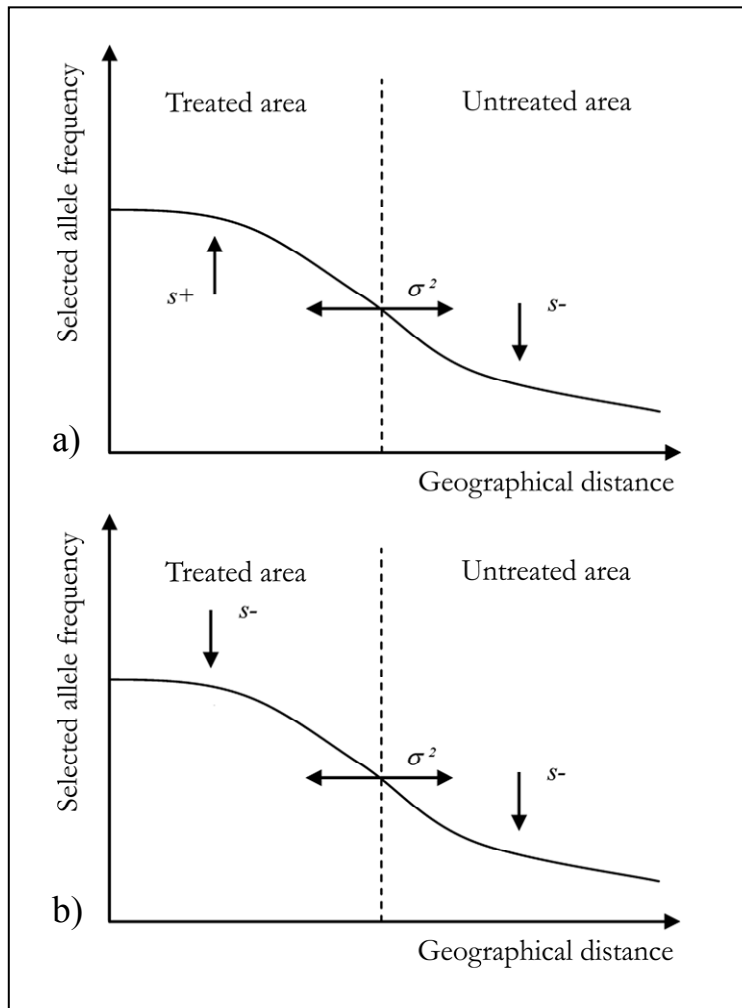


Figure 1. Schematization of the forces implicated in a migration selection cline: Positive ($s+$) and negative ($s-$) selection are generally considered as diversifying forces whereas migration between treated and untreated area, measured by the variance of parent-descendant dispersal distances (σ^2) is regarded as a homogenizing one.

a) Theoretical case in which both positive and negative selection operates on the cline at the time of the observation (potentially leading to a migration-selection equilibrium) => Expected situation before 2004 on the study site

b) Theoretical case in which positive selection at the origin of an expected cline is not operating anymore (potentially leading to a diminishing pattern of cline)
=> Expected situation after 2004 on the study site

MATERIALS AND METHODS

***Mycosphaerella fijiensis* in its environment**

M. fijiensis, the causal agent of the black leaf streak disease of banana (BLSD) is a haploid and heterothallic ascomycete fungus that spreads via three modes: the movement of infected plant material and the natural dispersal of spores produced either from sexual (ascospores) or asexual (conidia) reproduction (Gauhl 1994b). Differences exist between those two types of spores. For instance, asexual spores are produced during the early stages of leaf infection contrary to sexual ones (produced in late stages). Moreover, whereas conidia are supposed to be dispersed mainly over short distances, viable ascospores are wind-transported and may be carried over longer distances (Burt et al. 2003). *M. fijiensis* has been observed for the first time in Cameroon in 1980. It rapidly colonized all the banana plants of the country that can be divided in two contrasted farming systems: (i) the first one is the intensive cultivation of banana for exportation in wide industrial plantations that are under recurrent fungicide application, (ii) the second one is the extensive production of plantain banana as food crop in small plots or backyard gardens without any use of fungicides. A specific characteristic of the southwest Cameroonian banana producing area is the geographical coexistence of these two farming systems in a mosaic structure. With regard to BLSD control in commercial farms, a rational chemical control strategy based on 12–14 applications per year of several specific, highly curative and systemic fungicides has been established at the end of the 80' (Fouré, 1988). However, the local adaptation of the pathogen induced the emergence of high resistant strains levels to these fungicides. As a result, the chemical control strategy has progressively changed till the year 2004, date at which systemic fungicides have been replaced by unspecific contact protecting fungicides that prevent leaf infection through a multisite mode of action. Such contact fungicides reduce population density and prevent sexual reproduction of the pathogen but have not selected for any resistant strains yet. However, they must be applied each time a new banana leave rises (i.e., ~ every week) resulting in 40 – 50 applications per year.

Description of the experimental area and isolates sampling

The experimental site comprised a small industrial commercial plot (1 ha) located in SW Cameroon (Manjo, 4°48'N, 9°46'E) and banana plantain plots in its vicinity. This area was chosen mainly because the commercial plot was isolated from other industrial parcels and because many small plots of untreated plantain were continuously distributed in the surrounding. Moreover, this area was interesting because historical resistance levels data related to two systemic fungicides were available. Indeed, in October 2004, following a method based on in vitro spore ascospores germinative tube growth tests, monitoring surveys indicated that treated populations were characterized by around 60% and 25% of strains resistant to Strobilurins and Benzimidazoles chemicals respectively (Luc de Lapeyre de Bellaire, pers. com.).

Mycosphaerella fijiensis samples were collected in 2009 and 2010 from treated and untreated banana plant leaves infected with BLSD (Figure 2). Sampling sites were both situated inside and outside the treated plantation (up to 1700 m and 510 m from the border in 2009 & 2010

respectively) representing a total of 60 sites (30 in 2009 and 30 in 2010). At each sampling site, infected leaf fragments were collected from 3 closely spaced banana plants, and GPS coordinates were recorded with a precision of 3 m (WGS84 data format). 50 individual lesions per site were cut in the laboratory and dried for 1 day at 55°C before being stored at -20°C. Each individual lesion is considered as a single genotype.

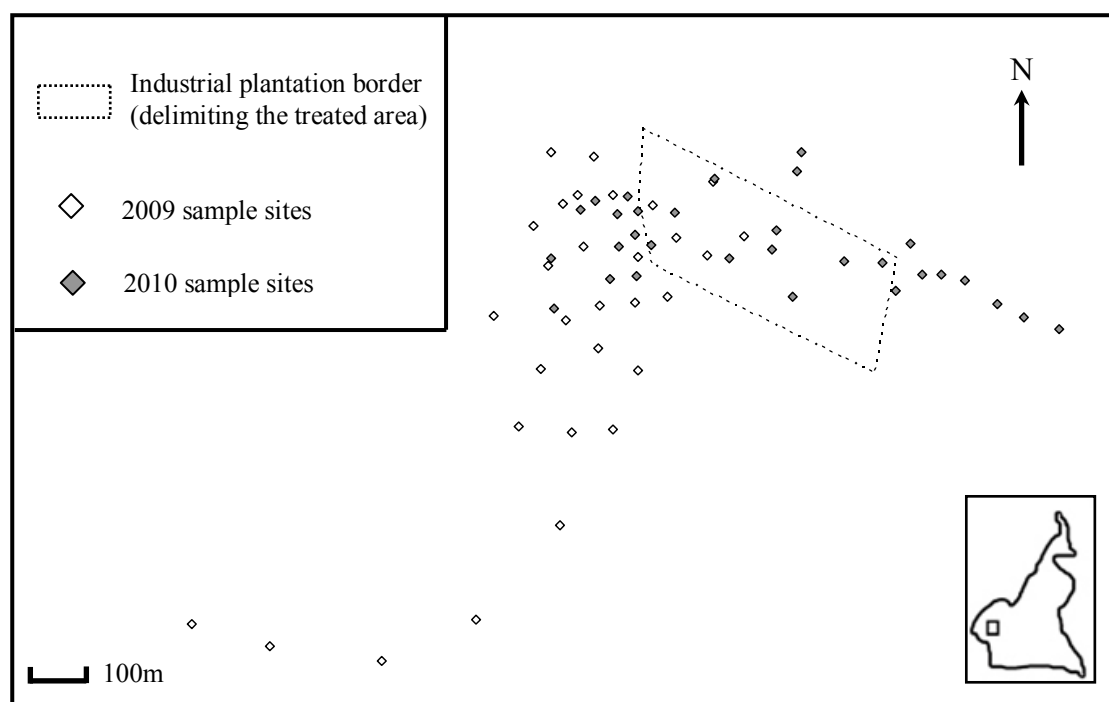


Figure 2. Sample site locations in the south-west Cameroonian area either inside or outside the industrial treated plantation considered (dashed line)

Molecular detection of fungicide resistance

Here, we investigated the local adaptation of *M. fijiensis* to two fungicides. The first family, the benzimidazoles, are antimetabolic products that inhibit tubulin polymerization. Resistance to benzimidazoles in *M. fijiensis* has been shown to be induced by a single point mutation at codon 198 of the β -tubulin gene: a nucleotide change of adenine instead of cytosine results in a modification of an amino acid in the beta tubulin (glutamic acid instead of alanine) (Canas-Gutierrez 2006). The second family, the strobilurins, are mitochondrial respiration inhibitor products that bind to the cytochrome b complex (QoI inhibitors). Resistance to strobilurins in *M. fijiensis* is also induced by a single point mutation in the mitochondrial cytochrome b gene leading to a change from glycine to alanine at position 143 (G143A) in the cytochrome b (Sierotzki et al., 2000). Even if most studies realized on benzimidazoles did not show any fitness cost of resistance, contradictory reports were made about the existence of pleiotropic fitness costs associated with both resistances in several studies focusing in other pathogenic fungi (McGee and Zuck 1981, Schuepp and Kung 1981, Ma et al. 2005, Fernandez-Ortuno et al. 2008). Such an assessment has never been

realized in *M. fijiensis* yet. To summarize the studied system, *M. fijiensis* resistance to Strobilurins and Benzimidazoles is coded by two independent loci with two alleles each.

DNA extraction were realised on 32 lesions/site using a procedure adapted to single dried lesions (more details in Robert et al. 2010). Two PCR/lesions were realised in order to amplify DNA fragment of 77 bp and 374 bp including both mutation points inducing resistance to Strobilurins and Benzimidazoles respectively. The detection of the mutation points was realised using the pyrosequencing technic at Bayer Crop Science laboratory. Resistance to benzimidazoles was monitored in 2009 sampling only whereas resistance to strobilurins was monitored both in 2009 and 2010 samplings.

Descriptive analysis

First, we used chi-square (χ^2) tests to look for significant differences between resistant alleles frequencies observed either between fungicides (Benzimidazoles vs. Strobilurins), areas (treated vs. untreated) or sampling time (2009 vs. 2010). We considered a level of significance $\alpha = 0.05$, and a degree of freedom $df = 1$ (corresponding to a critical χ^2 value of 3.84). The amount of linkage disequilibrium, measuring the non-random association of resistant alleles at two loci was computed by calculating the difference between observed and expected (assuming random distributions) allelic frequencies. This quantity could therefore be calculated only in sites in which the 4 alleles are present (resistant and sensitive alleles at both loci). We used the Gametic disequilibrium test implemented in GENEPOP (Rousset 2008) to test for significant values of linkage disequilibrium.

Second, we used the Cfit program version 7 (Gay et al. 2008) to fit genetic cline models using allelic count data. In this approach, loci are considered to have only 2 alleles, let say a_1 and a_2 and the frequency of a_1 can be fitted to decreasing function (the right allele at each locus should be chosen accordingly) as a function of geographical distance. Several clines can be fitted simultaneously to compare their slopes (concordance) and position (coincidence) and to construct complex hypotheses by constraining some of the parameters involved. Estimation is performed by maximum likelihood using a simulated annealing algorithm that includes Metropolis routines. We fitted data considering a scaled logit, characterized by four parameters:

$$h_1 + h_p(1 - h_1) \frac{e^{b(c-x)}}{1 + e^{b(c-x)}} \quad (1)$$

In this equation, h_1 is the lower asymptotic frequency, h_p determines h_2 (the higher asymptotic frequency) as $h_2 = h_1 + h_p(1 - h_1)$ (Figure 3), b is related to the maximum slope of the cline (the slope being $-h_p(1 - h_1)b/4$), and c is the centre of the cline. The width of the cline (w) can be calculated from these parameters as $w = (h_2 - h_1) / \text{slope}$.

Fits were first realised without constraining any parameter either in 2D (in this case, the angle maximising the slope of the cline is also estimated) or in 1D by taking into account the distance between each sites and the closest border of the industrial parcel. Second, we performed a likelihood search by successively constraining each parameter to be the same either between loci or

temporal sampling. Every fit was checked for convergence by reiterating the fit using different random seeds. Likelihood-ratio tests (LRTs) were used to compare the different models, and model choice was done considering a 5% significance level.

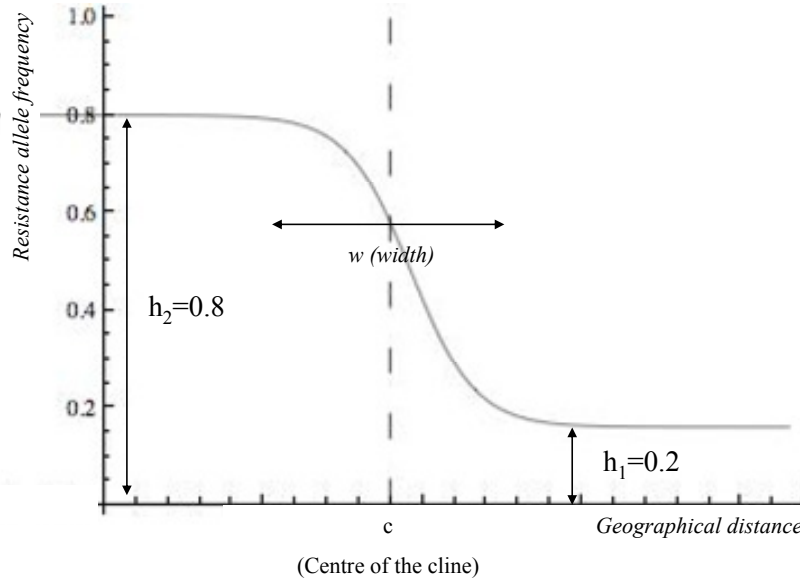


Figure 3. Schematization of the scaled logit function and its associated parameters. According to the equation given in the text, h_p would here be equal to 0.75 (which means that 75% of the remaining space between h_1 and 1 is occupied by the cline).

First estimates of fitness costs and migration rates from temporal differences in allelic frequencies

Estimation of fitness costs in absence of migration

We used a simple selection model introduced by Crow and Kimura (1970) (and reused in the fungicide specific case by Milgroom *et al.* (1989) to calculate the expected negative selection parameter associated with a decrease in the frequency of resistant strain through time in a isolated population. Let's denote x and y fungicide sensitive and resistant alleles respectively in frequencies $(1 - p)$ and p in the population. The evolution of the resistant allele frequency in this population through time can be written as:

$$\frac{dp}{dt} = (r_y - r_x)p(1 - p) \quad (2)$$

with r_x and r_y the absolute fitness of the two alleles (*i.e.*, their growth rate, Crow and Kimura 1970). The differential of selection s is given by $(r_x - r_y)$. This equation may be solved, so that the frequency p of the resistant allele at any time t can be expressed as a function of its frequency at a previous time 0 as:

$$\frac{p}{1 - p} = \frac{p_0}{1 - p_0} e^{st}. \quad (3)$$

Solving for s we obtain:

$$s = \frac{1}{t} \ln \frac{p(1-p_0)}{p_0(1-p)} \quad (4)$$

s quantifies the rate at which the frequency of resistance decrease through time due to fitness penalty in the absence of any positive selection.

Estimation of migration rate in absence of fitness cost

Here we aim at quantifying the intensity of migration from untreated to treated areas that would be sufficient to explain the decline in resistance allele frequency observed in the treated area. We assume that resistance-associated fitness costs are negligible. We further neglect spatial effects on migration (i.e., diffusion) and migration from treated to untreated areas. We consider non-overlapping generations.

Let's denote:

- N_T and N_{UT} the effective sizes of populations in treated and untreated areas respectively,
- and m the migration rate from untreated to treated areas.

Consider that the frequencies of resistant strains at time t are p_T and p_{UT} in treated and untreated areas. After migration, the treated area counts:

$$\beta N_T p_T + m N_{UT} p_{UT} \beta \text{ resistant individuals,}$$

for a total of:

$$\beta N_T + m N_{UT} \beta \text{ individuals.}$$

The frequency of resistance in treated areas therefore becomes:

$$p'_T = \frac{N_T p_T + m N_{UT} p_{UT}}{N_T + m N_{UT}} \quad (5)$$

In the untreated area, in absence of migration from the treated area, the frequency of resistant strains is stable

$$p'_{UT} = p_{UT}. \quad (6)$$

Equation (5) above can be solved, so that at any time t :

$$p_T = p_{UT} + \left(\frac{N_T}{N_T + m N_{UT}} \right)^t (p_{T0} - p_T), \quad (7)$$

where p_{T0} is the initial frequency of resistant strains in treated areas. This may be rewritten as:

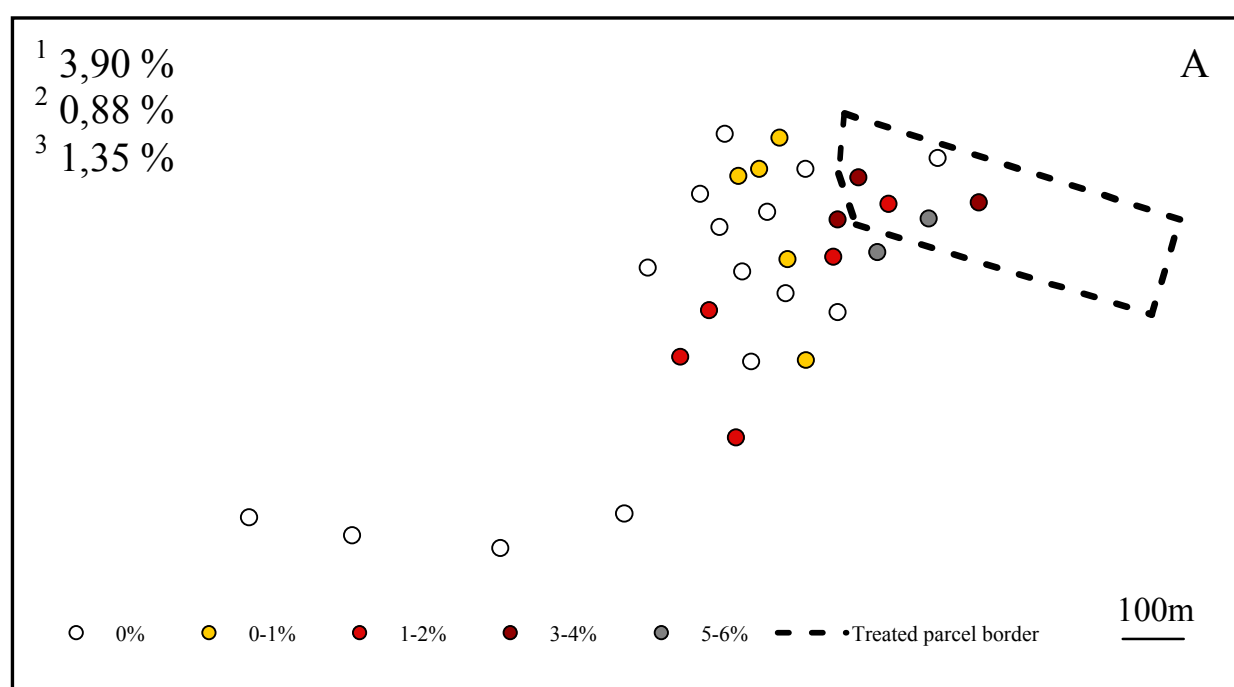
$$m \frac{N_{UT}}{N_T} = \left(\frac{p_{T0} - p_T}{p_T - p_{UT}} \right)^{\frac{1}{t}} - 1 \quad (8)$$

RESULTS

Global characterization of levels of resistance

In the 2009 data set, the frequency distribution of resistant alleles to Strobilurins and Benzimidazole were different (see Figure 4). The global frequency of resistance to Strobilurins was significantly higher than the one to Benzimidazoles ($\chi^2 = 5,38$), the latter appeared to be very low (1,35%). Both frequencies of resistant alleles were significantly higher in the treated than in the untreated area ($\chi^2 = 19,6$ and $6,3$ for Strobilurins and Benzimidazoles respectively). Out of the 25 untreated sites, 10 presented levels of Strobilurins resistance different from 0 with an average value of 5.60%. By another hand, no individuals displayed resistant alleles for both Strobilurins and Benzimidazoles (no cross-resistance). Linkage disequilibrium could be calculated on 6 sites only (among them, 4 sites were situated inside the treated area) but none of these values were significantly different from 0.

In 2010, the sampling and analysis strategy changed according to the results obtained in 2009. First, in 2010 we did not monitor resistance to Benzimidazoles because levels were very low the year before. Second, we did not sampled sites as far away from the parcel border but we explored other directions. We also increased the sampling in the close area around the border and inside the treated plantation (10 sites in 2010 instead of 5 sites in 2009) (Figure 4). The global level of resistance to Strobilurins was not significantly different between 2009 and 2010 ($\chi^2 = 1,43$). Finally, the frequency in resistant alleles was significantly higher in the treated area than in the untreated one ($\chi^2 = 14,3$). When comparing levels of resistance to Strobilurins between the two years, the decrease in frequency was significant inside the treated area ($\chi^2 = 5,79$) but the increase observed in the untreated area (probably due to differences in sampling strategy) was not significant ($\chi^2 = 3,06$).



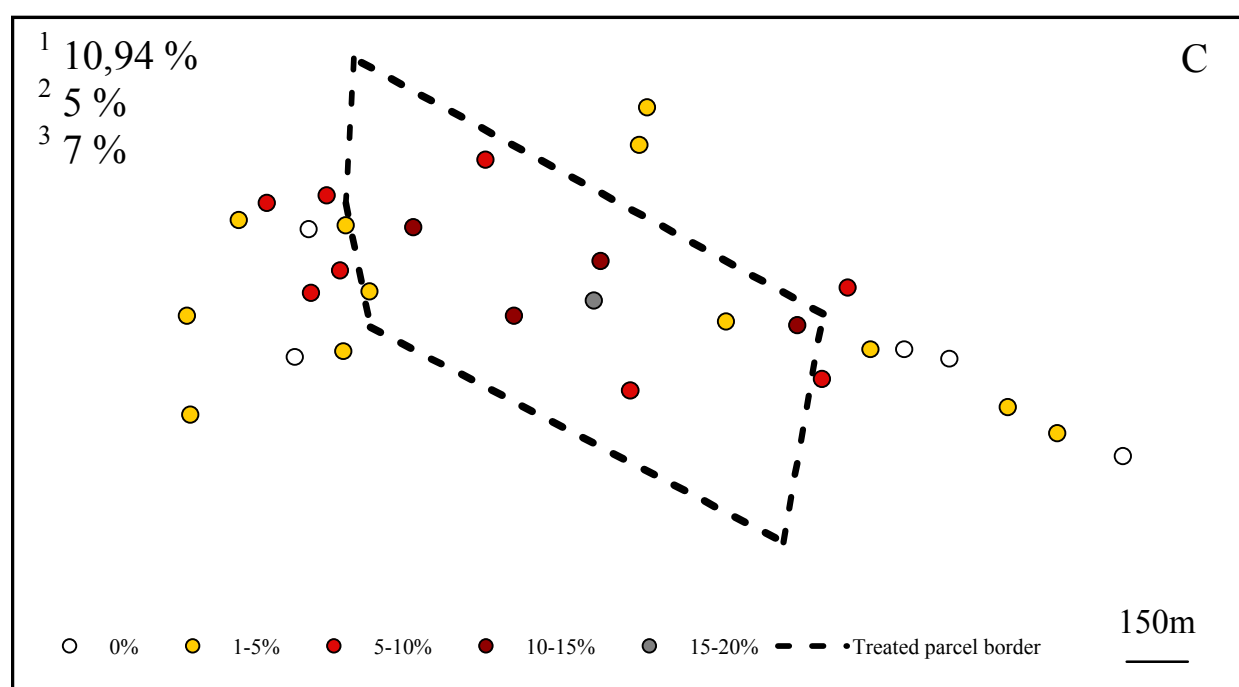
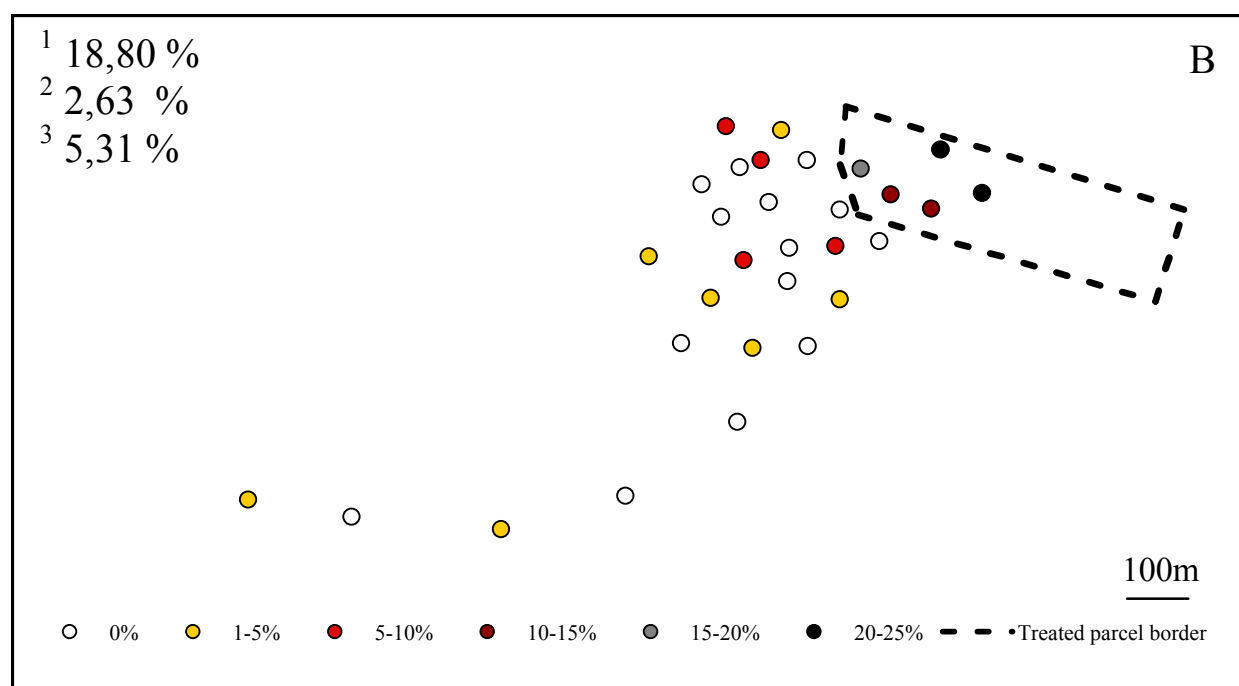


Figure 4. Frequency of the resistance allele measured at each sampled sites (N=33 ind. / site):

A – Benzimidazoles – Sampling 2009

B – Strobilurins – Sampling 2009

C – Strobilurins – Sampling 2010

Mean frequency of the resistant allele (in %) is given independently

● For all the sites (³)

● For sites either inside (¹) or outside (²) the treated area

Cline fitting analysis

Clines were fitted either in 2-dimensions (the angle of the cline maximising the slope is then inferred and coordinates are projected on this 1-D axe) or directly in 1-dimension by taking the distance of each sites to the closest border of the treated plantation geographical distance. The results were the same between these two procedures (results not shown). For the sake of clarity, we will present here only the results considering the distance to the closest border.

We realised two series of analyses. The first aimed at comparing the spatial frequency distribution of Strobilurins and Benzimidazoles resistant alleles in the first sampling (2009) whereas the objective of the second was to compare the clines of the Strobilurins resistant allele between the two temporal sampling dates.

Strobilurins vs. Benzimidazoles in 2009

First, the non-constrained models indicated that both loci presented a cline pattern (slope significantly different from 0) (not shown). Then, when successively operated constraints on the different parameters, we showed that parameter b was significantly larger for Benzimidazoles than for Strobilurins (indicating a significantly wider cline for Strobilurins). Moreover, upper and lower asymptotic limits (indicators of the frequency of resistant alleles in treated and untreated areas respectively) were significantly different between the clines at the two loci (Figure 6 and Table 1). Finally, the centres of both clines were not significantly different from each other and were situated at the border between treated and untreated areas.

Strobilurins in 2009 vs. 2010

Cline comparisons between both sampling dates in the case of resistance to Strobilurins indicate that the minimal frequencies of the resistant allele were not significantly different, whereas the maximal frequencies were significantly different between 2009 and 2010. This result is consistent with the one previously obtained and confirms that the decrease of the Strobilurins resistant allele frequency inside the treated area was significant. In 2010 the parameter b was not significantly smaller than in 2009 indicating that cline widths are not significantly different between the two years (Figure 7 and Table 2).

Preliminary estimates of fitness cost and migration rate from temporal differences in allelic frequencies

Fitness cost in absence of migration

From the difference in resistant allele frequencies between the historical value of 2004 and the measured value of 2009 in treated populations it was possible to infer fitness cost for both Benzimidazoles and Strobilurins. Between October 2004 and January 2009, there has been 50 months, which approximately corresponds to 100 generations of conidia (following the hypothesis that only conidia can participate from a generation to the next one in treated area). Using equation

4, we obtain a fitness cost of $s = -0.018$ per generation and $s = -0.020$ per generation for Strobilurins and Benzimidazoles respectively. Using the same approach, the fitness cost of strobilurin resistance could be computed between 2009 and 2010 (considering 24 generations). The result is $s = -0.026$.

Estimation of migration rate from untreated to treated areas in absence of fitness cost

Figure 8 shows the combinations of migration rate m and ratio of effective sizes N_{UT}/N_T that may account for the observed decrease of resistance frequency in treated areas using equation 8. We considered a number of generations of $t = 50$ for comparisons between 2004 and 2009 levels of both Benzimidazoles and Strobilurins resistant strains and $t = 12$ for comparison between 2009 and 2010 level of Strobilurins resistant strains (i.e. one generation/month following the hypothesis that ascospores only participate to gene flow between untreated and treated areas). It can be seen that for a realistic difference value between non-treated and treated population sizes ($N_{NT}/N_T=100$) (Luc de Lapeyre de Bellaire, pers. com.), small effective migration rates from untreated to treated areas (i.e., $m=0.4\%$, $m=0.3\%$ and $m=0.8\%$ for comparisons between 2004 and 2009 Benzimidazoles and Strobilurins resistant frequencies and 2009 and 2010 Strobilurins resistance frequencies respectively) are sufficient to account for the observed evolution of allele frequencies.

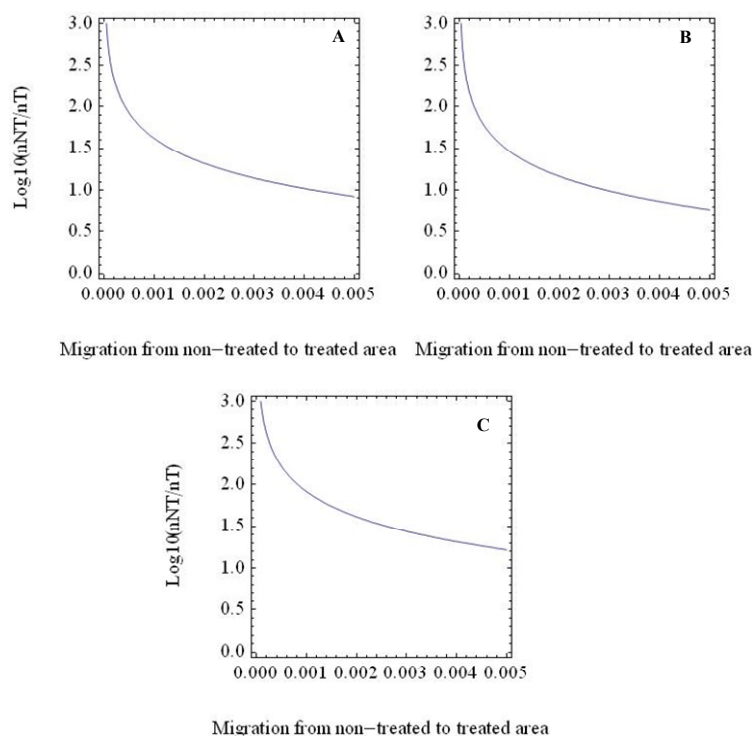


Figure 8. Plots of the estimated product between migration rate and effective sizes ratio between non-treated and treated populations (from equation 8).

We computed migration rate from non-treated to treated populations from the difference between:

A – 2004 historical level and 2009 observed level of benzimidazole resistant strains

B – 2004 historical level and 2009 observed level of strobilurins resistant strains

C – 2009 and 2010 observed level of strobilurins resistant strains (inside the industrial farm)

Note that the y-axis is in Log10 scale. For example, $\text{Log}_{10}(n_{NT}/n_T) = 2$ means that $n_{NT}/n_T = 10^2=100$ (effective size of untreated populations is one hundredfold that of treated populations)

resistant allele frequency

resistant allele frequency

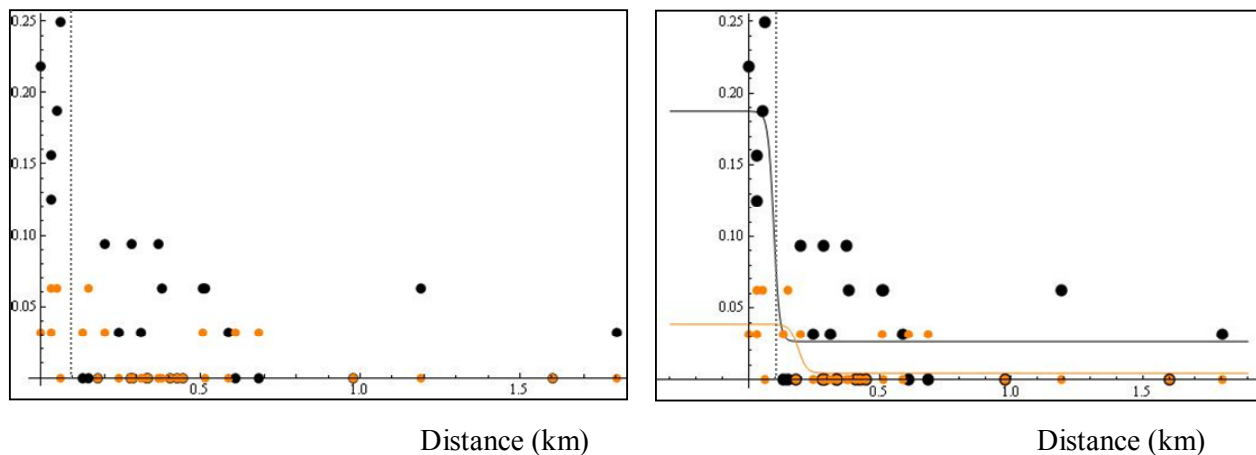


Figure 5: Comparison between the resistant allele frequency distribution to strobilurins (black) and benzimidazoles (orange) as a function of distance. For each site, the distance was calculated as the distance from the closest border of the treated parcel (dashed line). Points left to the dashed line corresponds to treated sites and points right to the dashed line corresponds to untreated sites. Fits obtained from unconstrained models were added on the right graph (see table n°1 for parameter values)

Model n°	Unconstrained models			Constrained models			
	1	2	1+2	3	4	5	6
	Strobi	Benzi	Strobi + Benzi	b	c	h1	hp
LL	-174.38	-62.58	-236.95	LL -238.97	-237.64	-242.71	-238.88
K	4	4	8	K 7	7	7	7
Centre	0.10	0.19		dev 4.04	1.39	11.51	3.86
b	811.17	62.27					
h1	0.03	0.00					
hp	0.17	0.03					
h2	0.19	0.04					
Width (m)	4.9	64.2					

Table 1. Cline pattern tests results for both strobilurins (strobi) and benzimidazoles (benzi) in 2009. Each cline was fitted using the scaled logit function. Log likelihood (LL) and the K estimated parameters are computed (Centre, b is related to the slope, h1 and hp indicates up and down trail heights). Note that the border is at the coordinate 0,1. Width (in meters) was calculated as $w = (h2 - h1) / \text{slope}$ considering that $\text{slope} = 1/4 b (-1 + h1) hp$.

For both loci, the slopes obtained from the non constrained model were significantly different from 0 (not shown). Constrained model were then fitted to test for significant difference in parameters value between strobilurins and benzimidazoles clines. In model 3, 4, 5, 6 the slope, centre, h1 and hp were constrained to be the same between strobilurins and benzimidazoles clines. Significant values of LRTs (5% level) are highlighted in bold (dev for deviance).

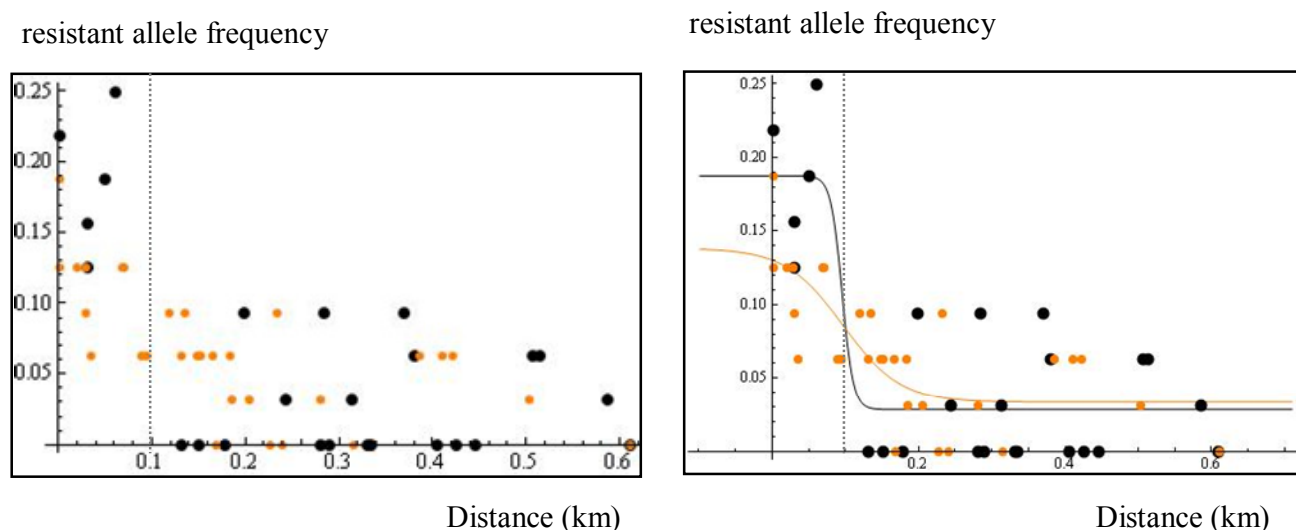


Figure 7. Comparison between the resistant allele frequency distribution to strobilurins in 2009 (black) and 2010 (orange) as a function of distance. For each site, the distance was calculated as the distance from the closest border of the treated parcel (dashed line). Points left to the dashed line corresponds to treated sites and points right to the dashed line corresponds to untreated sites. Fits obtained from unconstrained models were added on the right graph (see table n°2 for parameter values)

Model n°	Unconstrained models			Constrained models			
	1	2	1+2	3	4	5	6
	2009	2010	2009+2010	b	c	h1	hp
LL	-159.24	-226.11	-385.34	LL -385.89	-385.35	-385.42	-387.40
K	4	4	8	K 7	7	7	7
Centre	0.12	0.10		dev 1.09	0.02	0.16	4.11
b	112.45	25.01					
h1	0.03	0.03					
hp	0.16	0.11					
h2	0.19	0.14					
Width (m)	35.57	159.92					

Table 2. Cline pattern tests results for both 2009 and 2010 resistance to strobilurins. Each cline was fitted using the scaled logit function. Log likelihood (LL) and the K estimated parameters are computed (Centre, b is related to the slope, h1 and hp indicates up and down trail heights). Note that the border is at the coordinate 0,1. Width (in meters) was calculated as $w = (h2-h1)/\text{slope}$ considering that slope = $1/4 \cdot b \cdot (-1 + h1) \cdot hp$.

For both years, the slopes obtained from the non constrained model were significantly different from 0 (not shown). Constrained model were then fitted to test for significant difference in parameters value between 2009 and 2010 clines. In model 3, 4, 5, 6 the slope, centre, h1 and hp were constrained to be the same between 2009 and 2010 clines. Significant values of LRTs (5% level) are highlighted in bold (dev for deviance).

DISCUSSION

In this study, we monitored the frequency of resistance to Benzimidazoles and Strobilurins family fungicides in *Mycosphaerella fijiensis* at a small scale represented by a continuous habitat of treated and untreated populations. To our knowledge, this is the first study realised on a plant pathogenic fungus in which such a monitoring was realised. Indeed, previous studies focused on the monitoring of resistant alleles to several fungicides including both Strobilurins and Benzimidazoles either inside treated populations (reviewed in Ennos 1994) or inside populations after an interruption of treatments (Mc Gee & Zuck 1981, Chin et al. 2001, Avila-Adame & Köller 2003) in order to measure the evolution of resistant alleles frequencies and quantify selection (either positive or negative) parameters. Other studies included both treated and untreated populations (Dovas et al. 1976, Chen et al. 2007) but i) untreated populations were generally currently untreated at the moment of the study but previously treated and ii) the spatial scale considered was not continuous between those two environments. However, no resistance monitoring studies have been previously realised in a specific situation in which treated and “real” untreated populations may exchange individuals. We argue that it is probably caused by the rare nature of geographical proximity between treated and untreated populations in other pathosystems.

Statement of the forces expected to drive the evolution of resistant allele frequencies in the studied area

Since 2004 and neglecting genetic drift, resistant allelic frequencies are theoretically supposed to have decreased both in treated and untreated area because of a possible negative selection (i.e., fitness cost) and to be influenced by migration (that might carry sensitive or resistant strains between treated and untreated areas). However, since 2004, populations in the treated area are frequently sprayed with protectant fungicides that may influence the strength of migration in such theoretical expectations. Indeed, protectant fungicides that prevent the penetration of spores into leaves through a multisite mode of action are frequently applied resulting in a decrease of individual density in treated populations. Because *M. fijiensis* is self-incompatible (haploid heterothallic), the production of sexual spores (i.e., ascospores) requires important enough density for two sexually compatible lesions to spread and collapse on a same banana leave. Combined with other agricultural practices as pruning, the application of protectant fungicides impedes sexual reproduction in the treated area. Thus, the necrotic stage of the disease bearing the sexual structures of the fungus is very rare in treated plantations. As a result, the disease level is suspected to be maintained through the asexual production of conidia in the protected area. Because conidia, issued from asexual reproduction, are expected to disperse at short distances (i.e., to the nearby plants) (Gauhl 1994) whereas ascospores are actively projected in the air (Gauhl 1994, Yafetto et al. 2008) and wind-dispersed to higher distances (Burt et al. 2003), migration is expected to be strongly asymmetric from the untreated to the treated area.

Situation observed in the 2009 and 2010 samplings inside the treated area

A significant result of the study is that in January 2009, 50 months (i.e., approximately 50 sexual generations and 100 asexual generations for this pathogen) after the systemic fungicide application interruption, previously treated populations still presented detectable levels of resistance to Strobilurins and Benzimidazoles. First, the higher levels of both resistances in the treated area compared to the untreated area are consistent with the 2004 historical situation in which positive selection was operated only in the treated area. Second, the higher level of resistance detected in the treated population for Strobilurins in comparison to Benzimidazoles is also consistent with the one observed in historical data from 2004. As explained above, the occurrence of migration and/or negative selection between 2004 and 2009 may theoretically explain the observed decrease in resistant frequencies inside the treated plantation. Using simple theoretical considerations, we calculated values of both forces independently to explain the decrease in resistant alleles frequency in the treated area. We showed that weak negative selection (with inferred fitness penalty value of -0.018 and -0.020 per asexual generation for Strobilurins and Benzimidazoles resistant strains respectively) or weak effective migration rates from untreated to treated populations alone (0,4 ‰ and 0,3 ‰ assuming that effective size of untreated populations is one hundredfold that of treated populations) would be enough to explain the observed decrease in resistant frequencies inside the treated area.

The observation that the frequency of resistance to Strobilurins significantly kept on decreasing between 2009 and 2010 (i.e., 12 sexual or 24 asexual generations) in the treated area is also very interesting. We used the rate of decrease to infer a theoretical maximal expected fitness cost value (-0,026) and migration rate (0.8‰ in the case of an effective size ratio of 100) under the assumptions that both forces operated solely. The obtained values are consistent with those inferred between 2004 and 2009.

From the observation of the evolution of resistant frequencies in the treated populations only, both resistance associated fitness costs and unidirectional migration from untreated to treated areas, seem plausible explanations to the patterns observed. Indeed, pesticide resistance alleles are theoretically expected to exhibit fitness costs (Fisher 1930, Caspari 1952). However, even if fitness costs of resistance have never been studied yet on *M. fijiensis*, contradictory information has been reported for several pathogenic fungi. In theory, two general methods can be used to study the fitness costs of fungicide resistance alleles. The first involves detecting decreases in the frequencies of resistance alleles in isolated populations (i.e., without gene flow) that are not treated with fungicides for several generations. In the literature, such attempts have never showed any evidence of fitness penalties, either in the case of Strobilurins (Chin et al. 2001, Avila-Adame & Köller 2003) or Benzimidazoles (Dovas et al. 1976, McGee and Zuck 1981, Hanson 2006), but these monitoring have been conducted during a maximum of 6 generations. The second method involves comparing fitness components between resistant and susceptible individuals in the laboratory. For both fungicides families, inhibitory effects on mycelial growth (see Yan & Dickman 1996, Ma et al. 2003 for Benzimidazoles and Avila-Adame & Köller 2003, Markoglou et al. 2006 for Strobilurins)

have been detected. Further analyses are necessary to clarify those discrepancies in the case of *M. fijiensis*, notably by monitoring the field persistence of resistant alleles for longer periods.

From the present results, we cannot conclude on the relative strength of the forces implicated in the global decrease in resistant levels in the treated area. However, this preliminary analysis suggested that migration from untreated to treated populations could be implicated. This hypothesis is supported by observation of a gradient in disease severity from the border to the centre of treated parcels (data not shown).

Situation observed in the 2009 and 2010 samplings outside the treated area

In the 2009 and 2010, we also detected resistant strains at lower frequencies in untreated populations. Two contrasted arguments might explain this observation. First, despite their suspected low dispersal abilities, conidia migration from the treated area to the untreated area remains possible. Second, because of the small scale of the study area and the aerial nature of the fungicide applications, historical overflowed in systemic fungicide sprays could also be suspected.

When comparing of Strobilurins resistance levels outside the treated area between 2009 and 2010 we did not detect any significant difference. This result may indicate that there is either no fitness penalties associated with resistance to Strobilurins or fitness cost and gene flow from treated to untreated area. However, distinguishing between these two hypotheses was not possible from the present results. Further work is needed to clarify this point. For instance, the integration of migration in both directions into the theoretical model considered in the present study might help to determine if the expected migration rate necessary to counterbalanced the effect of any expected fitness cost is realistic or not.

PROSPECTS

We previously showed that, both in treated and untreated populations, selection and migration cannot be distinguished because they are expected to influence global pattern of resistant alleles frequencies the same way. However, differences in their spatial distribution are expected and genetic clines theory should provide a powerful statistical to disentangle their relative effects. Negative selection is supposed to induce a decrease in allelic frequencies everywhere with the same intensity whereas migration may generate different spatial patterns according to its spatial scale of occurrence. In a theoretical cline (Figure 1), negative selection may drop down the lower plateau whereas migration (if diffusion) is supposed to progressively decrease the slope at the centre of the cline. In the present study, we realised some preliminary descriptive cline analysis. We described the spatial pattern observed with a simple scaled logit cline that appeared to be adapted to fit such genetic data. Indeed, in 2009 two steep clines have been fitted to Strobilurins and Benzimidazoles spatial frequencies. The narrowness of the clines was consistent with the small level of gene flow between the two areas suspected in the present study. When comparing the 2010 cline with the 2009 cline, we detected i) a significant smaller higher plateau (characterizing levels of resistance in the treated area), no difference in the lower plateau (characterizing levels of resistance in the untreated area) and a smaller slope (and then a larger width) even if the difference was not significant

according the LRT. Further analyses are necessary to understand the relative importance of selection and migration on these clines. First, a new sampling was realised in April 2011 on the same sites as in 2010 and will be analysed soon. Second, during this recent sampling, the number of lesion and the surface of necrotic leaf tissues were recorded to quantify the difference in densities between treated and untreated areas. Third, we plan to build a theoretical model that integrates all the parameters implicated in the evolution of frequencies in this study case: negative selection, migration (that could either be diffusion or not), differences in effective sizes between treated and untreated populations and protecting efficiency of the contact chemicals. Comparing such a model with descriptive clines would be the only way to disentangle the relative effects of all the parameters on the evolution of frequencies and to infer their strength.

CONCLUSION

From a general point of view, the results obtained in this study allowed to better understand the dynamic of the resistant allele frequencies in this specific situation. First, it shows that resistant alleles may persist for many generations in populations not treated anymore with the fungicide in question that seem to indicate that if a fitness cost exists, it may be small. Second, it also illustrates that realistic migration rates between untreated plots and “treated with contact fungicides” populations could potentially induce a decrease of the resistance levels. We used the theoretical considerations presented in the present study to calculate the time needed to reach back a situation characterized by a critical level of resistance equal to 1% at which new forecasting strategies based on the limited use of systemic fungicides could be re-introduced at the scale of the studied area. We found that 60 and 35 generations of asexual reproduction (i.e., almost 1.5 and 2,5 years) would be needed for the Benzimidazoles and Strobilurins respectively. Finally our results also pointed out that past overflowed in fungicide applications might be implicated in the maintenance of significant levels of fungicide resistant strains at the outside border of treated plantations. Further work is needed on that point since it could have implications on the definition of specific resistance management strategies to be applied in such areas.

REFERENCES

- Agrios GN (2005) Plant Pathology, *Fourth Edition*.
- Avila-Adame C, Koller W (2003) Characterization of spontaneous mutants of *Magnaporthe grisea* expressing stable resistance to the Qo-inhibiting fungicide azoxystrobin. *Current Genetics* **42**, 332-338.
- Barton N, Hewitt G (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.
- Burt PJA (2003) Airborne dispersal of *Mycosphaerella fijiensis*. In: *Mycosphaerella Leaf Spot Diseases of Bananas: Present Status and Outlook. Proceedings of the Workshop on Mycosphaerella Leaf Spot Diseases, San José, Costa Rica*.

- Canas-Gutierrez GP, Patino LF, Rodriguez-Arango E, Arango R (2006) Molecular characterization of benomyl-resistant isolates of *Mycosphaerella fijiensis*, collected in colombia. *Journal of Phytopathology* **154**, 403-409.
- Caspari E (1952) Pleiotropic gene action. *Evolution* **6**, 1-18.
- Chen WJ, Delmotte F, Richard-Cervera S, *et al.* (2007) At least two origins of fungicide resistance in grapevine downy mildew Populations. *Applied and Environmental Microbiology* **73**, 5162-5172.
- Chin KM, Chavaillaz D, Kaesbohrer M, Staub T, Felsenstein FG (2001) Characterizing resistance risk of *Erysiphe graminis* f.sp tritici to strobilurins. *Crop Protection* **20**, 87-96.
- Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*.
- De Lapeyre de Bellaire L., Fouré E., Abadie C., J.. C (2010) Black leaf streak disease is challenging the banana industry. *Fruits* **65**, 327-342.
- Dovas C, Skylakakis G, Georgopoulos SG (1976) The Adaptability of the Benomyl-Resistant Population of *Cercospora beticola* in Northern Greece. *Phytopathology* **66**, 1452-1456.
- Endler JA (1977) Geographic variation, speciation, and clines. *Monographs in Population Biology* **10**, 1-246.
- Ennos RA, McConnell KC (1995) Using genetic-markers to investigate natural-selection in fungal populations
Canadian Journal of Botany-Revue Canadienne De Botanique **73**, S302-S310.
- Fernandez-Ortuno D, Tores JA, De Vicente A, Perez-Garcia A (2008) Mechanisms of resistance to Qol fungicides in phytopathogenic fungi. *International Microbiology* **11**, 1-9.
- Fisher RA (1930) *The genetical theory of natural selection*, Clarendon Press, Oxford. edn. Clarendon Press, Oxford. .
- Fisher RA (1937) The wave of advance of advantageous genes
Annals of Human Genetics **7**, 355-369.
- Fouré E (1988) Stratégies de lutte contre la cercosporiose noire des bananiers et des plantains provoquée par *Mycosphaerella fijiensis* Morelet. L'avertissement biologique au Cameroun. Evaluation des possibilités d'amélioration. *Fruits* **43**, 269-274.
- Gauhl F (1994) Epidemiology and ecology of black sigatoka (PhD Thesis), p. 120. INIBAP, Montpellier.
- Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* **62**, 2789-2806.
- Hanson L, Franc G, Panella L (2006) Characterization of genes associated with potential for fungicide resistance in *Cercospora beticola*. *American Society of Sugarbeet Technologists*.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**, 183-189.
- Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* **400**, 861-864.
- Lenormand T, Guillemaud T, Bourguet D, Raymond M (1998) Evaluating gene flow using selected markers: A case study. *Genetics* **149**, 1383-1392.
- Lenormand T, Raymond M (1998) Resistance management: the stable zone strategy. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 1985-1990.

- Lenormand T, Raymond M (2000) Analysis of clines with variable selection and variable migration. *American Naturalist* **155**, 70-82.
- Ma ZH, Michailides TJ (2005) Advances in understanding molecular mechanisms of fungicide resistance and molecular detection of resistant genotypes in phytopathogenic fungi. *Crop Protection* **24**, 853-863.
- Mallet J, Barton N, Lamas G, *et al.* (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in heliconius hybrid zones *Genetics* **124**, 921-936.
- Markoglou AN, Malandrakis AA, Vitoratos AG, Ziogas BN (2006) Characterization of laboratory mutants of *Botrytis cinerea* resistant to QoI fungicides. *European Journal of Plant Pathology* **115**, 149-162.
- McGee DC, Zuck MG (1981) Competition Between Benomyl-Resistant and Sensitive Strains of *Venturia inaequalis* on Apple Seedlings. *Phytopathology* **71**.
- Milgroom MG, Levin SA, Fry WE (1989) *Population genetics theory and fungicide resistance In Plant disease epidemiology* McGraw-Hill Publ, New York.
- Nagylaki T (1975) Conditions for the existence of clines. *Genetics*, 595-615.
- Robert S, Rieux A, Argout X, Carlier J, Zapater MF (2010) Optimized genotyping with microsatellite markers in the fungal banana pathogen *Mycosphaerella Fijiensis*. *American Journal of Botany* **97**, E130-E132.
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- Schuepp H, Kueng M (1981) Stability and tolerance to MBC in populations of *Botrytis cinerea* in vineyards of northern and eastern Switzerland. *Can. J. Plant Pathol* **3**.
- Sierotzki H, Parisi S, Steinfeld U, *et al.* (2000) Mode of resistance to respiration inhibitors at the cytochrome bc(1) enzyme complex of *Mycosphaerella fijiensis* field isolates. *Pest Management Science* **56**, 833-841.
- Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733-756.
- Yafetto L, Carroll L, Cui YL, *et al.* (2008) The Fastest Flights in Nature: High-Speed Spore Discharge Mechanisms among Fungi. *Plos One* **3**.
- Yan KY, Dickman MB (1996) Isolation of a beta-tubulin gene from *Fusarium moniliforme* that confers cold-sensitive benomyl resistance. *Applied and Environmental Microbiology* **62**, 3053-3056.
- Zhonghua M, Yoshimura MA, Michailides TJ (2003) Identification and characterization of benzimidazole resistance in *Monilinia fructicola* from stone fruit orchards in California. *Applied and Environmental Microbiology* **69**, 69 (12) 7145-7152.

Chapitre 3: Estimation directe d'un noyau de dispersion

Ce chapitre repose sur la réalisation d'une expérimentation présentée dans l'article 5 (Spore dispersal in a plant pathogenic fungus: on the estimation of anisotropy and individual dispersal functions from one-generation disease gradients) visant à mesurer de façon directe la distribution des distances de dispersion (i.e., le noyau de dispersion) des spores sexuées et asexuées de *Mycosphaerella fijiensis*.

Il existe deux grandes classes de méthodes directes permettant de mesurer la distribution des distances de dispersion (Bullock et al. 2006). La première, dite « Lagrangienne », consiste à suivre dans le temps la trajectoire d'un individu donné alors que la seconde, dite « Eulérienne » repose sur la détermination du nombre d'individus arrivant en un point donné de l'espace. Dans le cas des champignons phytopathogènes, les spores sont de trop petite taille (i.e., non visibles à l'oeil nu) et la méthode Lagrangienne n'est pas applicable. En revanche, l'utilisation d'une méthode Eulérienne est envisageable et son principe général s'organise en deux points clefs. La première étape consiste à capturer des spores en plusieurs points de l'espace. Pour cela, il est possible d'utiliser des pièges physiques ou dans certaines situations des pièges biologiques (par exemple, le comptage de symptômes sur du matériel vivant constitue une façon d'estimer le nombre de spores arrivant en ce point). L'avantage, du piège biologique est qu'il permet de mesurer une dispersion efficace (i.e., flux de gènes) ce qui n'est pas le cas avec les pièges physiques. La seconde étape consiste à calculer la distance parcourue par les différentes spores capturées, ce qui nécessite de pouvoir retrouver *a posteriori* leurs origines géographiques. Cette étape implique généralement d'utiliser une source artificielle d'inoculum marquée et localisée ainsi que d'empêcher la succession et la juxtaposition de plusieurs cycles de dispersion à partir de cette source.

Puisqu'il est rarement envisageable (pour des raisons techniques et logistiques) de réaliser des captures en tout point de l'espace, le choix du nombre et de la répartition spatiale des pièges est une étape importante dans la réalisation des approches Eulériennes. Elle l'est d'autant plus qu'elle conditionne la fastidieuse tâche de détection d'évènements rares de dispersion (souvent à longues distances). En effet, puisque la dispersion est généralement limitée dans l'espace, la probabilité d'observer un évènement de dispersion diminue avec la distance géographique à une source donnée. De plus, en deux dimensions, la surface potentielle sur laquelle une spore peut atterrir augmente intrinsèquement avec la distance à la source (par exemple, à 10 mètres d'une source la surface potentielle de la zone réceptrice (πr^2) est de $s \sim 80\text{m}^2$ alors qu'à 50 mètres elle est déjà de $s \sim 1900\text{m}^2$). Ainsi, afin d'augmenter la probabilité d'observer un évènement rare de dispersion à longue distance, il est généralement recommandé d'augmenter i) la taille des pièges et ii) le nombre de directions échantillonnées lorsque la distance à la source augmente. La capture de spores dans plusieurs directions est également indispensable afin de pouvoir déterminer si le processus de dispersion est de nature anisotrope (i.e., dépendant de la direction) ou pas. Cependant, considérant un nombre donné de pièges, il existe un grand nombre de possibilités concernant leur répartition spatiale

(aléatoire, en grille, en secteurs...) et la nature du dispositif permettant une caractérisation optimale de la distribution des distances de dispersion sera différente selon celle du noyau de dispersion (ce qui est problématique puisque c'est justement ce noyau que l'on cherche à estimer en réalisant l'expérimentation). Dans ce contexte, il est possible de tester de façon théorique la capacité de plusieurs dispositifs expérimentaux à permettre une estimation la plus complète et robuste possible de la distribution des distances de dispersion en fonction de différents noyaux (Skarpaas et al. 2005). Ainsi, une telle approche est importante à considérer lors de la définition d'un dispositif expérimental visant à être mis en place en conditions naturelles, notamment lorsque des informations *a priori* sur la nature du processus de dispersion sont disponibles.

Une fois que le nombre d'évènements de dispersion à différentes distances et dans différentes directions a été mesuré, l'utilisation d'une approche statistique permettant de synthétiser ces observations est généralement nécessaire. Cette étape a pour but de décrire la forme de la décroissance des évènements de dispersion avec la distance ainsi que de détecter une éventuelle anisotropie. Dans le cas de spores dispersées par le vent, la propriété anisotropique du processus de dispersion est fréquente et peut se manifester de deux façons différentes. On distingue l'anisotropie de densité (lorsque le nombre moyen de spores observées est fonction de la direction) de l'anisotropie de distance (lorsque la distance moyenne de dispersion des spores est fonction de la direction). Ces deux types d'anisotropie peuvent parfois exister dans des directions différentes (Soubeyrand et al. 2007). L'analyse statistique consistant en l'ajustement des données observées à des fonctions théoriques doit donc explicitement prendre en compte ces différents facteurs.

Ainsi, la mesure expérimentale de la distribution des distances de dispersion requiert énormément de précautions, aussi bien dans la mise en place du dispositif que dans l'analyse statistique des données. L'expérimentation présentée dans ce chapitre a été réalisée au Cameroun en octobre-novembre 2010 et est l'aboutissement de plus d'une année de préparation notamment pour la mise en place du dispositif et la préparation de la source de dispersion. Nous avons essayé de prendre le maximum de précautions possibles afin de limiter les risques de contamination durant l'expérimentation. Les résultats présentés dans cet article sont issus d'analyses statistiques récemment réalisées. Ces analyses demeurent préliminaires et devront être approfondies. Des pistes d'analyses supplémentaires seront d'ailleurs proposées en fin du manuscrit. Il m'a paru important de bien détailler les différentes étapes de l'expérimentation (préparation + réalisation) ainsi que de l'analyse statistique au sein de ce manuscrit. En conséquence, son format (long) devra probablement être réduit en simplifiant certaines parties (à définir en fonction du journal ciblé pour la soumission).

Spore dispersal in a plant pathogenic fungus: estimation of anisotropy and individual dispersal functions from one-generation disease gradients

A. RIEUX^{*}, J. CARLIER^{*}, F. BONNOT^{*}, E. KLEIN[¶], S. SOUBEYRAND[¶], V. RAVIGNE^{*}, J. ESSOH
NGANDO[‡], A. MEHL[†] and L. DE LAPEYRE DE BELLAIRE[•]

^{*} CIRAD, UMR BGPI, Campus international de Baillarguet – Montpellier, France

[¶] INRA, UR 546, Biostatistiques et Processus Spatiaux – 84914 Avignon, France

[‡] Centre Africain de Recherches sur Bananiers et Plantains (CARBAP) – Njombe, Cameroon

[†] Bayer CropScience AG Alfred-Nobel-Str. 50, Building 6240 D-40789 Monheim, Germany

[•] CIRAD, Persyst, UPR Syst. Banan. Ananas, TA B-26 / PS4 – Montpellier, France

SUMMARY

Given their biological significance, determining both the shape and the tail of the dispersal distances distribution of plant pathogens is essential. Here, we report two field experiments carried out to independently measure disease gradients caused by sexual and asexual spores of the wind-dispersed banana plant pathogen fungus *Mycosphaerella fijiensis*. To avoid contamination from either external or secondary foci, we used a traceable source of inoculum composed of fungicide-resistant strains in an environment where those strains were not present, and the experiment was conducted over a single generation of the pathogen. Disease gradients were assessed over a high number of plant trap plots installed up to 1000 m along eight directions radiating from the source. We analysed data by simultaneously testing for anisotropy and adjusting two theoretical kernels differing in the shape of their tails. Contrasting results were obtained between ascospore and conidia dispersal gradients. On the ascospore experimental site, we characterized both a steep gradient in the first few metres of each direction and rare long-distance dispersal (LDD) events up to 1000 m from the source in two of the eight directions. As a result, the ascospore gradient appeared to be better fitted to a heavy-tailed distribution. Our results also indicated that although there was no significant preference in the direction taken by ascospores, they dispersed on average further in two different directions that were not correlated with wind patterns. On the conidia experimental plot, very few dispersal events occurred outside of the central plot. However, we observed a gradient up to 12.5 m from the source in one of the eight directions. Here, the disease gradient was better fitted by the exponential kernel, and anisotropy in both density and distance appeared to be correlated with averaged daily wind gust. Our results, which provide the first estimation of dispersal kernels for this plant pathogenic fungus, should be helpful in better predicting both the speed of disease expansion in new areas and the diffusion of emerging genotypes such as fungicide-resistant strains under different management strategies.

INTRODUCTION

Understanding dispersal processes is of particular interest for at least two reasons. First, as dispersal is directly related to the colonisation of species in new environments, it determines the speed of spatial expansion (Skellam 1951) and the resulting spatial genetic structure (Austerlitz 1997). Second, dispersal determines the level of gene exchange between implanted populations, and their evolutionary potential (Lenormand 2002). In plant pathogenic fungi, a better knowledge of these processes may improve our capacity to predict patterns of disease spread both at local and global scale (Shaw 1995), and to define new efficient management strategies for disease control (Lenormand & Raymond 1998, Forster & Gilligan 2007). However, for many species, and especially for wind-borne pathogens, dispersal is expected to be leptokurtic (Lacey 1996, Wingen et al. 2007), *i.e.*, there is an excess of short- and long-distance dispersal (LDD) events in comparison to the values expected under a Gaussian distribution. As a result, dispersal gradients of wind-dispersed pathogens are typically very steep, with most propagules travelling only a short distance from the source of inoculum (Gregory 1968). However, the relatively few propagules travelling far from the source (*i.e.*, LDD) are crucial to disease expansion in new territories (Scherin 1996, Brown & Hovmoller 2002, Sackett et al. 2005) and for the diffusion of emerging genotypes such as fungicide-resistant strains (Gilligan 2008).

Three main approaches can be used to study wind-dispersal of fungal propagules. The first is mechanistic (*i.e.*, physical models) and includes fine and realistic modeling of processes such as air-release mechanisms and the aerology of propagules (Tufto et al. 1997, Soubeyrand et al. 2008, Burie et al. 2011). However, despite significant advances, these approaches cannot be used routinely because of the still imperfect understanding of the underlying mechanisms and the “cost” of numerical simulations. The second approach aims to estimate gene flow from genetic data (*i.e.*, indirect approaches) through measures of genetic differentiation among populations or individuals (see Broquet et al. 2009 for a complete review). Indirect methods generally allow the estimation of a single synthetic parameter (*e.g.*, an average dispersal distance per generation), but do not enable a better representation of all dispersal events, including LDD, because they provide a measure of gene flow that is averaged over time and space (but see Austerlitz et al. 2004, Robledo-Arnuncio et al. 2006, Saccherri et al. 2008). The third approach is based on the direct (*i.e.*, real-time) measurement of dispersal distances and can be achieved by determining how many propagules from an identified source disperse at specific locations (Bullock et al. 2006). In theory, such direct methods allow the estimation of the complete distribution of dispersal distances through statistical inference of an individual dispersal function (IDF) or kernel (Lavigne et al. 1998, Klein et al. 2006, Darmency et al. 2009). A thorough estimation of the whole shape of the IDF is necessary because characterisation of the tail is crucial to quantifying the intensity of LDD (Klein et al. 2006).

Nevertheless, despite all these available methods, LDD is very difficult to characterize (Nathan 2001), and specific precautions should be considered in any experimental design seeking accurate IDF estimation (Bullock et al. 2006). First, the size of the experimental site and the intensity of the source of inoculum are often critical, since both have to be big enough to provide measurable disease levels at long distances. Second, because spore dispersal might be anisotropic (*i.e.*,

directionally dependent) in windborne fungi (Gregory 1968), intensive trapping should be arranged in several directions around the source (Bullock et al. 2000, Soubeyrand 2007). The trap sampling design and the respective size and shape of the source and trap plots is very important and needs to consider several factors (Bullock et al. 2006, see Skarpaas et al. 2005 for an optimizing approach using Monte Carlo simulations). Finally, since individual spores cannot be tracked due to their small size, it is crucial to be highly confident in the fact that all observations (spores trapped, or disease symptoms) made at different sampling sites are due to dispersal events from the localised inoculum source. This requires both an easily identifiable source of inoculum so that external contamination can be excluded, and the avoidance of secondary foci of dispersal by performing the experiment during only a single generation of the pathogen.

Here, we were interested in measuring the IDF of *Mycosphaerella fijiensis* - the causal agent of black Sigatoka (Black Leaf Streak Disease, BLSD) of banana and plantain. This ascomycetous fungus is known to be a major cause of yield loss, and is thought to induce ecological perturbation because of the massive use of fungicide required for its control (Churchill et al. 2010, de Lapeyre de Bellaire et al. 2010). A specific feature of this fungus is its natural spread via the simultaneous and continuous dispersal of spores produced either from sexual (ascospores) or asexual (conidia) reproduction. Conidia are produced in younger stages of the disease, *i.e.* from stage 1 to stage 4 according to Fouré's scale (Fouré 1987), while ascospores are produced in necrotic spots at the latest stage of the disease, *i.e.* stage 6. Air release mechanisms of both spore types are influenced mainly by climatic conditions (Carlier et al., 2000 in Jones, 2000) but are noticeably different. Indeed, conidia are lightly attached to conidiophores and laid on the leaf surface so that they are blown off passively by wind or water to be dispersed, whereas ascospores are expelled actively into the air from the perithecia during wet periods before being wind dispersed (Gauhl 1994). Some previous studies used rotorod spore traps to clarify the relative importance of ascospores vs conidia dispersal inside and outside plantations in a mountainous area of Costa Rica (Burt et al. 1997, Rutter et al. 1998, Burt 2003) and in the banana production area of the Atlantic coast (Gauhl 1994). These studies showed that both types of spores are present in the air within a banana plantation. However, reports on the prevalence of conidia vs. ascospores in the two locations are contradictory, with conidia being more frequent inside plantations located in mountainous areas. On the other hand, ascospores were more frequent when traps were localised outside banana plantations, supporting the hypothesis that ascospores might be carried over long distances whereas conidia are dispersed mainly over short distances or at the plant level (Burt et al. 1998). Given this context, other studies were carried out with the objective of measuring dispersal distances through spore trapping or disease gradient analysis but none allowed an accurate IDF to be determined. Effectively, in these studies, either external contamination was impossible to avoid (Burt et al. 1998, Abadie et al., *in preparation*) or disease gradients were observed over several disease cycles (Rutter et al. 1998; Amil et al., 2007). For instance, Abadie et al. (*in prep*) analysed disease gradient patterns at the scale of a small parcel (30 x 30m) implanted in an area characterised by low disease pressure and during one single generation of the pathogen. The result of this latter study showed the presence of a sharp and quick decrease in the number of lesions, suggesting an average dispersal distance of less than 30 m. However, LDD events could not be detected because of the small size of the sampling area, and occurrence of external contamination was also suspected. More recently,

Amil et al. (2007) used an inoculum source composed of strains resistant to a specific fungicide in order to track their spread over several kilometres during a six-month period. Weak levels of resistance were detected up to 6 km away from the source but the succession of several pathogen life cycles made differentiation of contamination events caused by dispersal from the primary source or from secondary foci impossible.

The main objective of the present study was to estimate a dispersal kernel in *Mycosphaerella fijiensis* under natural conditions, including the detection of rare LDD events. Because it has been previously argued that ascospores and conidia might exhibit different patterns of dispersal, we decided to measure two independent IDFs for both sexual and asexual spores. For this purpose, two experimental plots were specifically set up in Cameroon in order to measure BLSD gradients i) up to 1000 m in several directions from an artificial and marked inoculum source, and ii) during a single generation of the pathogen. We analysed data by testing simultaneously for anisotropy and adjusting several theoretical kernels differing in the shape of their tails.

MATERIALS AND METHODS

General principle of the experiment

This experiment was designed to measure the dispersal kernel of both sexual and asexual spores of *M. fijiensis* by analysing disease gradients at different distances from a specific marked source of inoculum after one single generation of the pathogen. This source of inoculum was composed of strains resistant to a specific fungicide placed in the center of the two experimental sites. All banana trees sampled at different distances from the source were sprayed with the fungicide to which the inoculum source was resistant. Strains present around the experimental plot are thought to be fully sensitive, thus the only inoculum theoretically capable of causing any symptoms on the banana plants in the experiment is the source at the centre of the experimental plots.

Production of resistant inoculum source

In this study we used a strobilurin-resistant inoculum as a dispersal marker. Strobilurins (also called QoIs) are systemic fungicides that inhibit mitochondrial respiration of fungi by binding to the cytochrome bc1 enzyme complex (complex III) at the Qo site (Bartlett et al., 2002). To obtain an inoculum source highly resistant to these fungicides, 50 banana plants were sprayed every week, from January to September 2010, with Bankit 25 SC at the rate of 100 g azoxystrobin/ha in 20 l of water and using 1‰ Triton X 100 as a wetting agent. These banana plants were located in an area 50 km away from the experimental site to avoid contamination from this source. The frequency of evolution of resistant strains was monitored at different periods as described below. The frequency of strains resistant to strobilurin increased from 0% in January 2010 to 94% in October 2010 (Figure 1), and the last samples collected the day before the experiment displayed very high levels of resistance (96% for lesions bearing conidia and 98% for lesions bearing ascospores). Two different sources of inoculum were prepared the day before the start of the experiment. For the source composed of ascospores, leaves in position 7 were cut from 50 banana plants. These leaves

carried a large amount of necrotic tissue with stage-6 disease, where ascospores are produced and not conidia. This necrotic tissue was stapled to artificial leaves made of steel and mosquito net (Figure 2). We produced 48 artificial leaves (80 cm long and 40 cm wide) and all were installed on 5 artificial trunks (plastic pipe), which allowed us to concentrate a large surface area of necrotic tissue (12 m² in total) into a small soil surface area (Figure 2) with 5 artificial banana trees.

For the source composed of conidia, we selected 5 banana plants on which all necrotic tissue with a disease level above stage 4 had been removed, so that the remaining leaves bore lesions exclusively of stages 1 to 3, which produce only conidia. These 5 banana plants were uprooted and transplanted to the experimental plot (Figure 2).

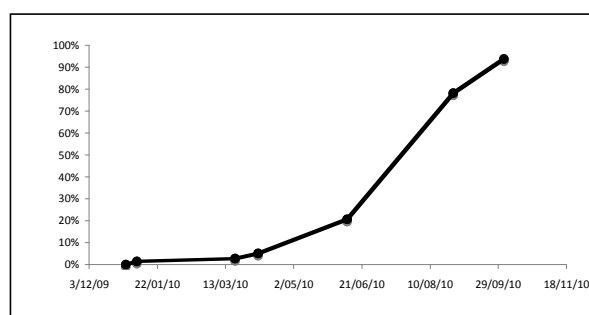


Figure 1. Evolution of the QoI- resistant allele in the banana plot where the source of inoculum has been produced. Fungicide treatments with azoxystrobin were realized once a week during 9 months at a frequency of 1 spray/week (see text for details)

Experimental estimation of the intensity of inoculum sources

Four days before installation of the inoculum sources, leaf samples were collected on 10 of the 50 banana plants sprayed with azoxystrobin.

Source of ascospores. Necrotic leaf samples (stage 6) on leaves in position 7 were collected. On each leaf sample collected, two leaf discs of 2.2 cm in diameter were cut with a pastry cutter and stapled to a filter paper disc 90 mm in diameter (5 necrotic discs per filter paper). The filter papers were immersed in distilled water for 5 min and then placed in the lid of a Petri dish agar plate for two hours. After ascospore discharge, the filter papers were removed and, after 48 h, the number of ascospores per disc was counted under a microscope. The average value of this number was taken as an estimate of the number of ascospores produced per surface of necrotic area in the source (the total surface of the ascospore source was 12 m²).

Source of conidia. Leaf samples bearing young symptoms (stage 2) of BLSD were collected from leaves in position 4. On each leaf sample collected, 1 leaf disc of 2.2 cm in diameter was cut with a pastry cutter in a place where lesions were sufficiently dense (about 5 lesions/cm²), and the number of lesions/disc was counted. In order to remove all conidia from the lesions, discs were applied successively 6 times to a Petri dish agar plate. The number of conidia/disc was counted 48 h later under a microscope. The number of conidia produced per lesion was then calculated from this value (from lesion counting we estimated that each banana tree of the source bore 1,000 lesions at stage 2).

Experimental design

The study was conducted in a large industrial palm tree and hevea plantation (SOCAPALM Company) in southwest Cameroon (4°14'N, 9°57'E) in October-November 2010. This area seemed an ideal location for our experiment for several reasons. First, this industrial plantation allowed the experiment to be implemented over a large area without any banana plants, thus limiting gene flow from external sources. Second, industrial banana plantations are situated far from this site, which should prevent any incidence of strobilurin-resistant strains from an external source (banana plantations where resistant strains have been reported are located more than 35 km away). Third, the experiment was located in a large (300 ha), recently planted (2009) parcel of hevea. The area was flat, with no obvious obstacles to dispersal (*e.g.* forest or hedges). Moreover, the young hevea trees were thin and also did not represent a physical barrier to spore dispersal. Banana vitroplants were planted in June 2010 according to the two experimental designs outlined above within this large planting of young hevea. The banana plants belonged to the cultivar Grande Naine (*Musa acuminata*, triploid AAA, Cavendish subgroup) - a variety highly susceptible to BLSD - and were obtained from a nursery implanted far from any commercial plantation where fungicide resistance to strobilurins has been reported.

Ascospore experimental site. Banana plants were planted at different distances, from 4 to 1000 m, and on each of eight directions (indexed A, B, C, D, E, F, G, H) radiating from a 4x4 m central square (Figure 3). In order to measure precisely the expected short-distance gradient, we constructed a tight planting of single banana plants (*i.e.* natural traps) up to 104 m from the source. At further distances, the intensity of sampling sites was progressively relaxed but the number of banana plants/site increased with distance in order to increase the reception surface and thus the probability of observing rare events at such distances (Figure 4). In some directions, the dimensions of the hevea plot did not allow the installation of banana trap plants at the longest distances (see Figure 3). A total of 511 banana plants were planted on 192 sites. We also planted 4 vitroplants inside the 4x4 m central square to obtain data on the number of spores produced by the source of inoculum that do not disperse outside the central square.

Conidia experimental site. Banana vitroplants were planted in an open area without hevea trees. This area was located 2 km away from the closest part of the ascospore experimental plot (axis F), and 3 km from the ascospore inoculum source. Because we hypothesised that conidia disperse much less far than ascospores, the size of this experimental site was smaller, and one banana plant was planted every 2.5 m from a central point up to a distance of 25 m in each of the eight directions (indexed a, b, c, d, e, f, g, h) (Figure 3).

Finally, the GPS coordinates of all sites were recorded (WGS84 data format). From June to October, standard cultural practices including fertilization were applied to each banana plant. In October, most of the banana plants were in good physiological condition (very few died) in the ascospore site. However, the situation was quite different on the conidia site where many plants died or did not grow due to bad soil conditions (Figure 3). Indeed, it was noted that the conidia experimental site did not correspond to a cultivated area but was over a possibly compacted abandoned airstrip.

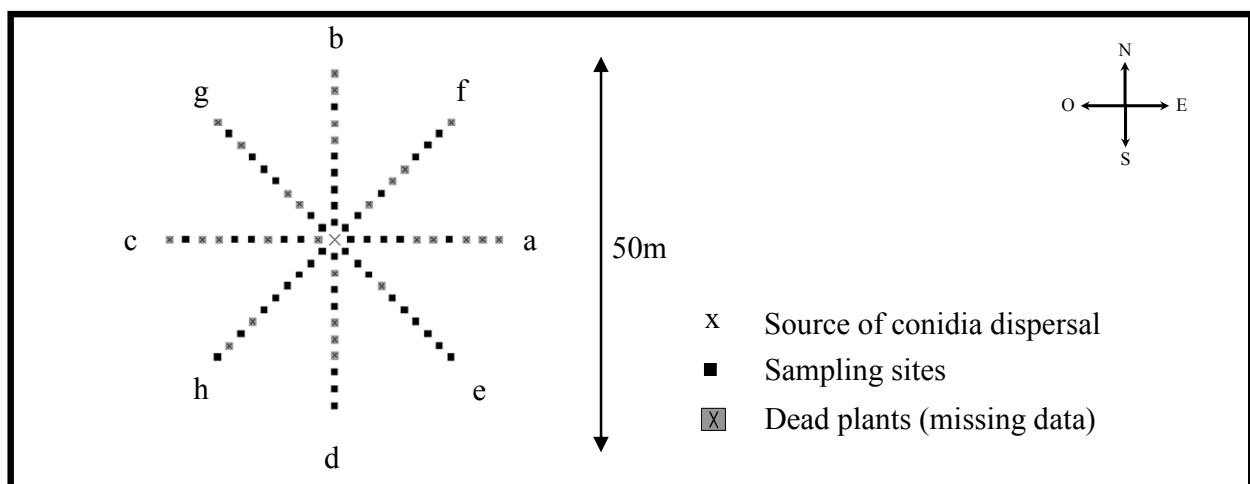
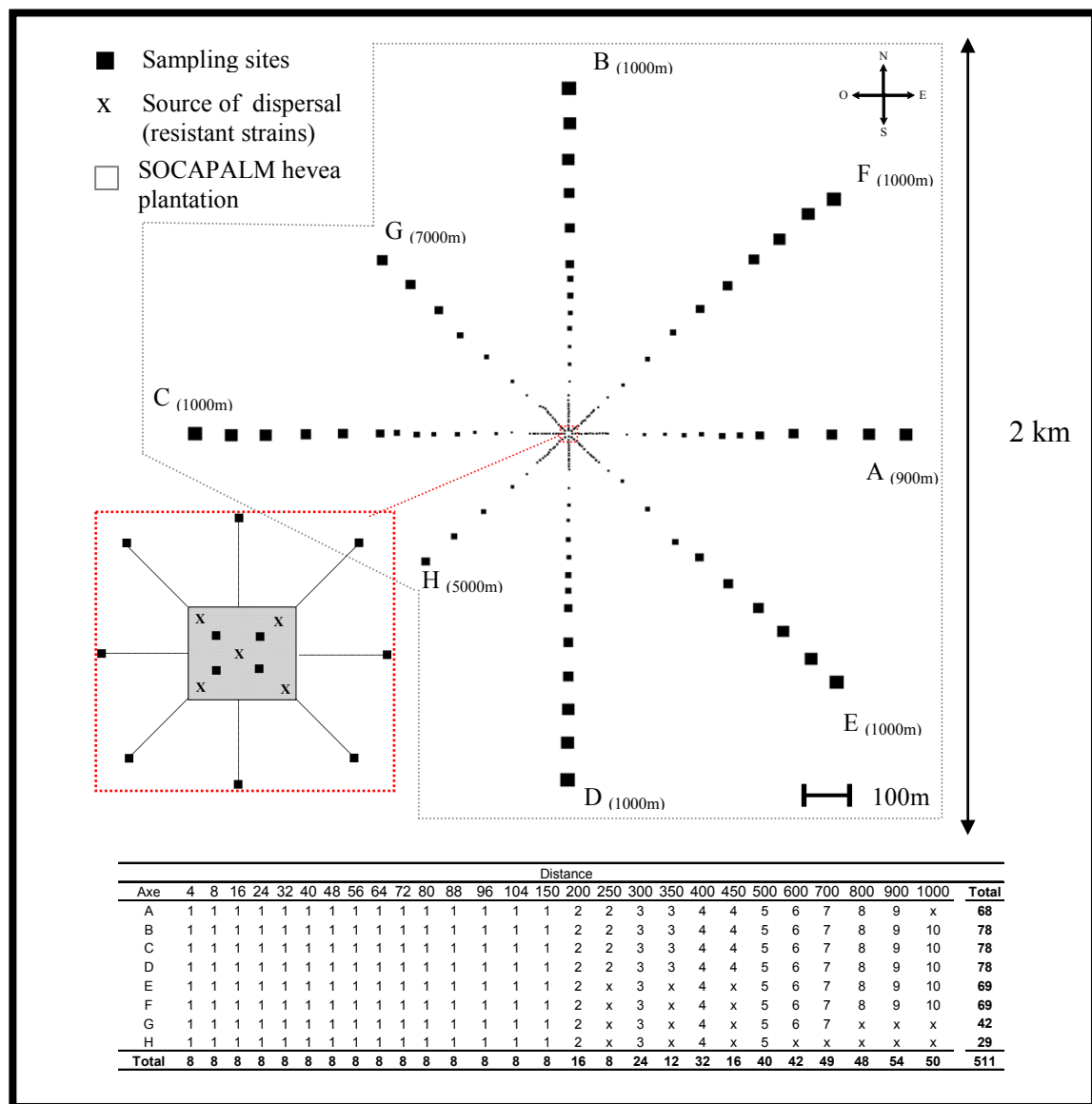


Figure 3. Plots layout of both ascospores (up) and conidia (bottom) experimental designs implanted in Cameroon. For the ascospores dispositive, the sizes of the squares (schematizing sampling sites) are proportional to the reception leaf surface and the number into brackets indicates for each axe the distance between the centre and the most distant sampling site. Number of natural traps (i.e., banana plants) at each site are indicated in the table.



Figure 2. Pictures of the resistant source of inoculums used on the two experimental plots. The two pictures at the top represent the artificial source of ascospores and the two pictures at the bottom illustrate the natural source of conidia.

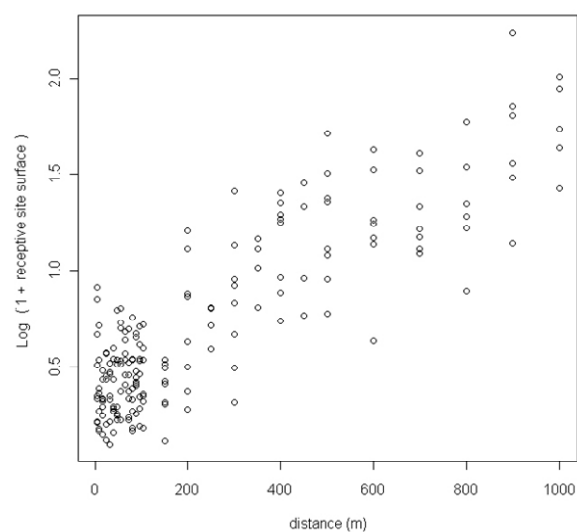


Figure 4. Ascospores experimental site: the receptive surface at the different sites (calculated as the sum of the contaminated and observed leaf surface of each banana plants) increases as a function of geographical distance.

Installation of inoculum sources

Four days before the installation of the inoculum sources, the banana plants were sprayed with Bankit 25 SC as described above. The day before the installation of the inoculum sources (D_{-1} , 4th October), the leaves of all banana plants were removed from the experimental site and extracted, except the last unfurled leaf, called the cigar, and the first leaf under the cigar (F1), *i.e.* the leaves in which primary infections mainly occur. This pruning was carried out to eliminate all inoculum source inside the experimental site, since no ascospores or conidia are produced on these young leaves. On the 5th of October (D_0), the inoculum sources (48 artificial leaves with necrotic spots and the 5 banana plants without necrotic spots) were transported carefully to the experimental site and were installed at the center of each of the two experimental plots. On the ascospore plot, we installed five artificial banana plants in a 4x4 m central square (see Figures 2 & 3). On the conidia plot, we planted 5 banana plants in a central area of 1.5x1.5 m.

Measuring disease gradient after source installation

After the installation of inoculum sources, banana plants were sprayed with Bankit 25 SC every week, as described above, in order to prevent further leaf contamination by external inoculum sources composed of susceptible strains (Figure 5). One week after D_0 , the leaves of banana plants located close to the source began to be monitored every 3 days in order to detect the first symptoms caused by the source inoculum. Leaf 1 present at D_{-1} was removed at that time, and only the leaf that was unfurled at D_0 was observed further. This leaf was exposed to the source during the period during which contaminations occur. After detection of the first symptoms (D_{0+1}), the inoculum source was removed. The period D_{0+1} was considered the incubation period, but since conidia might be formed at these early stages, we also considered this period as the latent period. Then, in order to avoid assessing lesions from a second generation of spores, a final counting of lesions was made on these leaves at time $2 \times (D_{0+1}) - 4$. Only stage 2 and stage 3 lesions were considered since stage 1 lesions can be confused with other leaf injuries. The length (L) and width (l) of each leaf were also measured in order to calculate leaf surface area ($L \times l \times 0.83$).

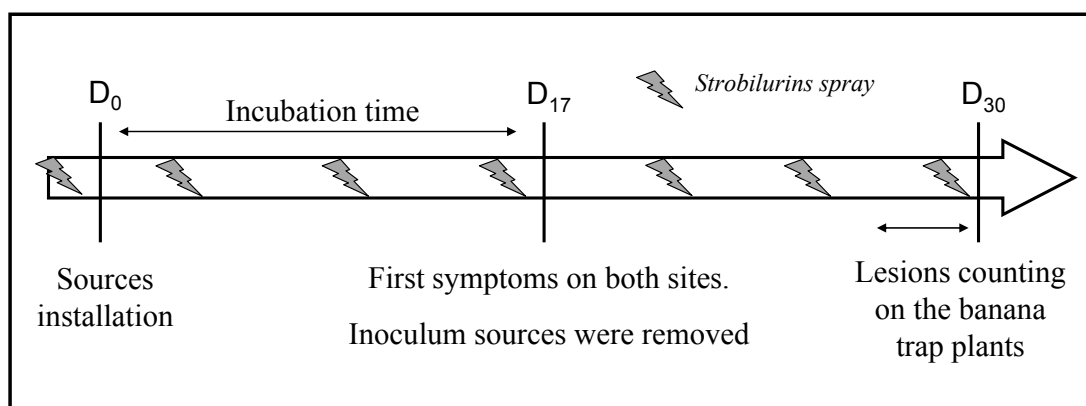


Figure 5. Schematization of the main steps realized during the experiment.

Meteorological data

We used a WS2800 (La crosse ® Technology) weather station to record wind gust and speed, temperature, and humidity at a frequency of one measurement point every 15 minutes during the entire duration of the experiment. Daily rainfall was recorded using a rain gauge.

Sampling and molecular detection of strains resistant to strobilurins

Molecular detection of strobilurin resistance was performed at three steps during the experiment:

(i) On the banana plants prepared for use as source of inoculum

Seven temporal samplings were taken to monitor the evolution of strobilurin resistance between January and October 2010. Each sampling was carried out on 10 banana plants chosen randomly inside the treated parcel, and was composed of 30 lesions / plant.

(ii) Inside the experimental plot and on banana plants around the experimental plot before the installation of the inoculum source

One banana plant at the extremity of each axis, one banana plant in the center of the experimental site, and 15 banana plants surrounding the experimental plots were sampled before installation of the source of inoculum; 30 lesions were sampled on each of the selected plants.

(iii) On the leaves of banana plants where lesions were counted in the experimental plots

After lesion counting, individual lesions were sampled at each site with a maximum sampling size of 32 lesions. For sites in which several banana plants were present (sites > 150 m from the source), lesions were sampled randomly on every plant.

All samples were analysed according to the same protocol. For each lesion, a standardized fragment (2 mm in diameter) was cut in the laboratory and dried for 1 day at 55°C before being stored at – 20°C. A maximum of 10 lesion fragments from the same plant was bulked for DNA extraction according to the procedure detailed in Robert et al. (2010). In *M. fijiensis*, QoI resistance is conferred by a single mutation in the mitochondrial cytochrome *b* (cyt *b*) gene (Sierotzki et al., 2000). This mutation leads to an amino acid change from glycine to alanine at codon 143 (G143A) of the cyt *b* gene. A 37-bp DNA fragment including the G143A mutation was amplified by PCR. The mutation point frequency for each sample of *M. fijiensis* was detected using a pyrosequencing technique at Bayer Crop Science Laboratory.

Data notation

Each site was indexed by the combination of a letter (indicating the direction, see Figure 3) and a number (representing the distance in metres from the source). For each site, we noted TNL [the total

number of lesions (field-counted)], and TDL [the total density of lesions (in number of lesions/leaf m²)]. After estimating the percentage of resistance in the bulked lesions, the TNL was transformed to give the total number of resistant lesions (TNRL) as $TNRL = (FR) * TNL$, where FR = frequency of resistance in the bulk sample. Finally, the total density of resistant lesions (TDRL) (in number of resistant lesions/leaf m²) was obtained from the TNRL.

Statistical analyses

Dispersal is anisotropic if the behaviour of the dispersed propagules depends on the direction. This angular dependence can occur in various ways. Here, we focus on two sorts of dependence represented by two independent summary functions characterizing dispersal in density and distance, respectively (Soubeyrand et al. 2007). The first describes the direction taken by the spores during transit, whereas the second provides the expected distance between the spore's source and the location of its deposition, given the direction taken by the spore. We used von Mises function (Fisher 1995) to model those two anisotropy functions. Note that von Mises function is a unimodal function (*i.e.*, there is only one main dispersal direction) that is not constrained by any prerequisite model.

Modeling of the anisotropy in density

Let f denote the von Mises directional density function, reflecting the probability that a spore is deposited in an increasing width sector with angle $d\varphi$ radiating from the source in direction φ (with φ in $[0^\circ-360^\circ[$ (Figure 6.A).

$$f(\varphi) = \{2\pi I_0(\sigma)\}^{-1} \exp\{\sigma \cos(\varphi - \mu)\} \quad \{1\}$$

$$I_0(\sigma) = (2\pi)^{-1} \int_0^{2\pi} \exp\{\sigma \cos(\theta - \mu)\} d\theta \quad \{2\}$$

Parameter μ is the expected dispersal direction, and σ represents the variability of dispersal directions around μ . For instance, $\sigma = 0$ means that dispersal directions are distributed uniformly in $[0^\circ-360^\circ[$, and $\sigma = \infty$ means that dispersal directions are all equal to μ (see example in Figure 6.B). Note that I_0 is an integrative constant.

Modeling of the anisotropy in distance

Let g denote the mean distance function and be proportional to a von Mises function. The function g describes, for any direction φ in $[0^\circ-360^\circ]$, how far from the source a spore is deposited on average. More precisely, it indicates the mean dispersal distance of spores deposited over an increasing width sector radiating from the source in direction φ (Figure 6A).

$$g(\varphi) = g_0 \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\varphi - \nu)\} \quad \{3\}$$

In which ν is the direction in which the mean dispersal distance is the larger and k controls the variability of the expected dispersal distance as the dispersal direction varies. For instance, $k=0$ means that dispersal distances are the same whatever the dispersal direction. Note that g_0 is a multiplicative constant measuring how far from the source spores are deposited (*i.e.*, it is not a distance).

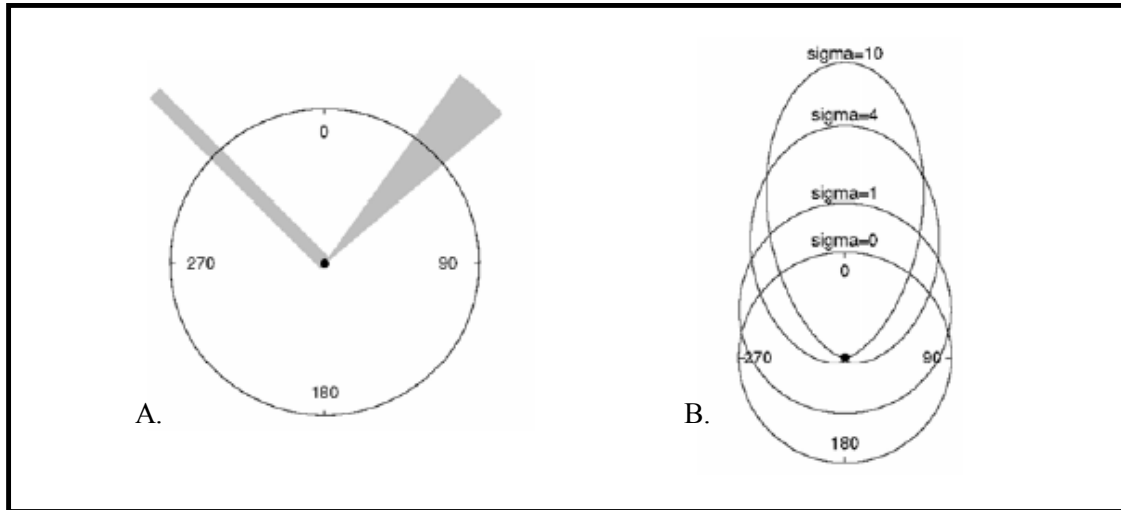


Figure 6. A- Graphical representations of two sorts of plots radiating from the source location (black dot) used to interpret the anisotropy functions. Left gray shape: constant width sector; Right gray shape: increasing width sector. B- The von Mises functions for different values of the parameter sigma (σ) which controls the variability around the mean 0° . The von Mises function is an unimodal function.

Mixing anisotropy functions with IDFs

First, an underlying model should be defined to describe the number of lesions by leaf. For any leaf located at x , the number of lesions is assumed to follow a negative binomial distribution (a generalization of the Poisson distribution) with mean $S(x)$, where $S(x)$ (called infectious potential at x) measures the expected quantity of spores that are deposited at location x that can, potentially, infect a leaf.

The infectious potential at location x is generated by the sum of the 5 artificial sources ($y_{i,i} = 1, 5$) located in the centre of the parcel, and integrates the anisotropy functions f and g as:

$$S(x) = \sum_{i=1}^{i=5} S(x, y_i) \quad \{4\}$$

with

$$S(x, y_i) = s_0 \frac{f(\varphi)}{g(\varphi)^2} \exp \left\{ \frac{\rho}{g(\varphi)} \right\} \text{ for the exponential kernel} \quad \{5\}$$

$$S(x, y_i) = s_0 \frac{f(\varphi)(b-1)(b-2)}{g(\varphi)^2} \left(\frac{g(\varphi)}{1+\rho} \right)^b \text{ for the geometric kernel} \quad \{6\}$$

Where (φ, ρ) are the polar coordinates of x . Note that s_0 represents the sum of the five source strengths (considering each to be equivalent). Source strength is defined here as the number of spores released by the source and travelling out of the source area. The parameter b is a shape parameter than can take a minimum value of 2 (decreasing values lead to heavier tails). The quantity $S(x) dx/s_0$ is the probability that a spore dispersed by the source is deposited in the infinitesimal rectangular area dx centered at x . The infectious potential function could be seen as a two-dimensional version of the individual dispersal function that takes anisotropy functions into account.

Parameter estimation and tests of significance

In the model, there are 6 or 7 parameters: the source strength s_0 , the direction parameters μ and σ , and the distance parameters v , k and g_0 . In the case of the geometric kernel, there is a 7th parameter b describing the shape of the tail. All parameters were estimated simultaneously by maximum likelihood using an adaptive barrier algorithm (Rohatgi 2003). Likelihood ratio tests (LRTs) were performed to see i) whether the directional density function is uniform ($\sigma = 0$) or not ($\sigma > 0$), and ii) whether the mean distance function is constant over $[0^\circ-360^\circ]$ ($k = 0$) or not ($k > 0$).

Mean distance travelled by spores

We computed two types of mean distance travelled by spores. The first, obtained directly from the inferred infectious potential function, is averaged over the 8 directions and takes into account the anisotropy parameters. The second was calculated from the raw data (*i.e.*, without fixing any kernel), and on each direction independently as the non-central moment of order 1. By definition,

the non-central moment of order 1 is: $\frac{R}{N}$, R being the total distance travelled by the N spores. To

obtain an independent value for each of the 8 directions, we divided the total surface S divided into $k=8$ elements, the k_{th} element being of surface s_k , of density f_k , at distance r_k . Then the number of spores on s_k is $n_k = f_k s_k$ and the total distance travelled by the n_k spores is $R_k = n_k r_k = f_k s_k r_k$. The

average distance is thus obtained by dividing $\frac{R}{N}$ as :
$$\frac{\sum_{k=1}^K r_k f_k s_k}{\sum_{k=1}^K f_k s_k} \quad \{7\}$$

Correlation between density, distance and wind patterns.

We used the Spearman rank correlation coefficient (and the associated two-sided p-value) to test for significant correlation between spore density or distance travelled by spores and wind patterns.

RESULTS

Potential source intensity

From experimental measurements we observed a mean value of 2410 ascospores expelled per leaf disc, and 36 conidia per lesion, resulting in an average potential source intensity of 7×10^6 ascospores and 1.5×10^5 conidia.

Disease assessment

In both experimental sites, the first symptoms were observed at D_{17} on the plants closest to the source. We thus considered that the maximum incubation time (IT), as well as the maximum duration of the latency period (LP), under the conditions of this experiment was 17 days (it could be shorter if the first ascospores or conidia were released after D_0). Disease assessment (*i.e.*, lesion counting) was then conducted between days D_{26} and D_{30} , *i.e.* a period of less than $2 \times LP$.

Ascospore experimental site

In the ascospore plot, lesions were detected in 168 of the 192 trap sites (87.5%), and the TNL observed on each site ranged between 0 (24 sites) and 4356 (1 site located inside the source square). A sharp decrease in the disease gradient was observed over the first 100 m in each of the eight sampling directions (see Figure 7A). In some directions, (axes C, D, E), the TDL increased at greater distances from the source (between 800 and 1000m), suggesting external contamination from banana plants present in the close vicinity of the experimental site in these directions.

Conidia experimental site

Many data were missing from the conidia plot because of the high mortality rate of the banana plants (29/80 plants, *i.e.*, 36%). On the remaining banana plants, we observed lesions on only 16 plants (31%). The TNL observed ranged between 0 (35 sites) and 73. No BLSD symptoms were detected at a distance greater than 12.5 m from the central source (Figure 8A).

Molecular detection of resistance

None of the several lesions or necrotic tissue sampled before the start of the experiment either inside or outside the experimental site displayed the presence of any single strobilurin-resistant strain, confirming the absence of the strobilurin-resistance allele in the surrounding wild populations.

On the ascospore experimental site, the percentage of the resistance allele detected varied from 0% (mainly for sites situated at extreme distances, close to the borders of the experimental site) to 100%, with an average value of 75%. This result (which is illustrated by the difference between Figure 7A and 7B) indicated i) that the artificial inoculum source was efficient, and ii) that despite the fungicide treatment, some sensitive spores could induce lesions on sprayed leaves. We observed

resistant lesions up to 1000 m from the source (at two sites, TDRL = 1.8 at C₁₀₀₀ and TDRL = 5.6 at D₁₀₀₀) suggesting the occurrence of rare LDD events at these sites.

On the conidia experimental site, the percentage of resistance detected varied from 0% (at two sites) to 100%, with an average value of 71%.

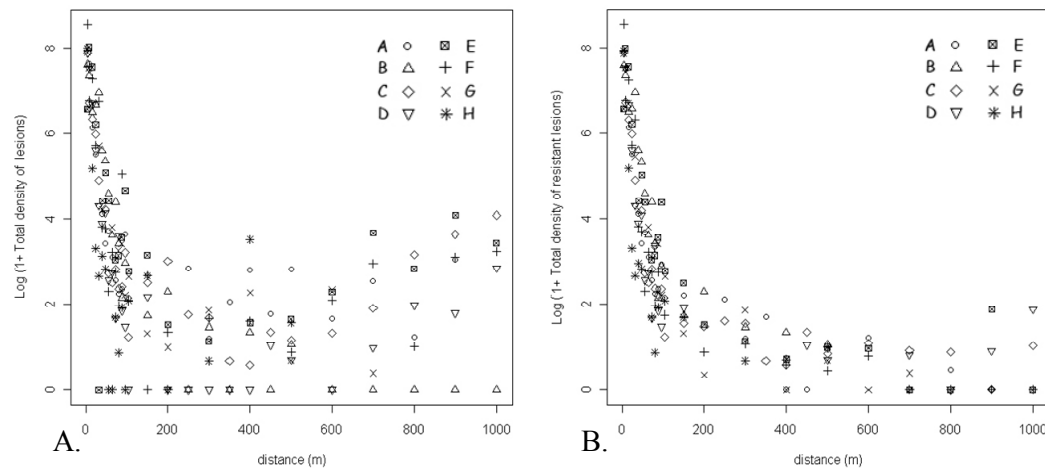


Figure 7. Ascopores experimental site: Total density of lesions (i.e. TDL, A) and total density of resistant lesions (i.e. TDRL, B) as a function of geographical distances from the central source. The different axis of the experimental design are represented separately. TDL and TDRL are in “log (1+the value)” scale.

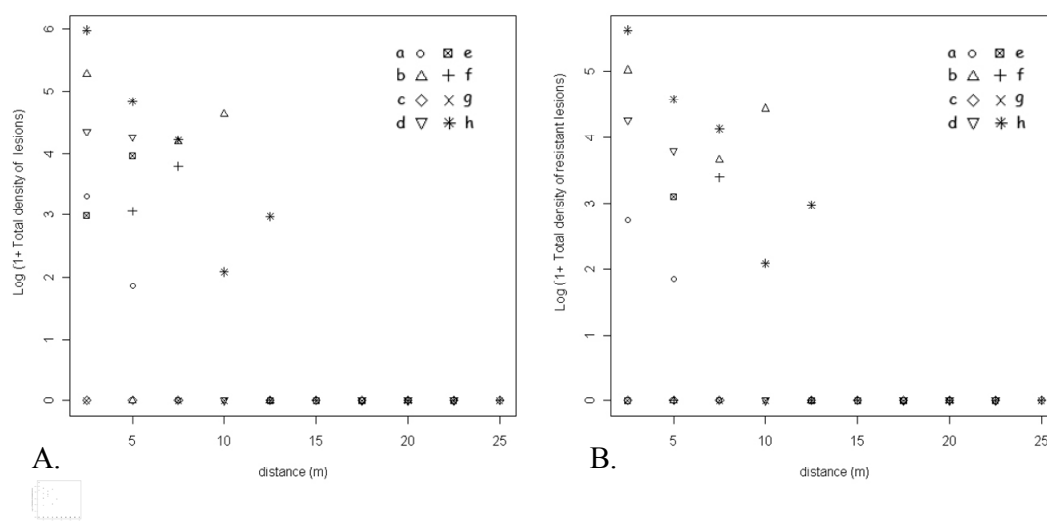


Figure 8. Conidia experimental site: Total density of lesions (i.e. TDL, A) and total density of resistant lesions (i.e. TDRL, B) as a function of geographical distances from the central source. The different axis of the experimental design are represented separately. TDL and TDRL are in “log (1+the value)” scale.

Weather report

The cumulative wind speed (sum of wind speeds over all measurement points) and the wind frequency computed for each measured direction is given in Figure 9 for the period where the inoculum sources were installed in the experimental designs (D0-D17). Figure 9A shows that the highest cumulative wind speed averaged over this period was in the SW direction (*i.e.*, blowing from the NE). However, we observed the opposite pattern when considering only data recorded between 5 and 9 a.m. (Figure 9B) with highest cumulative wind speed in the NE and the N. This time interval has been described previously as potentially the main ascospore discharge period in *M. fijiensis* because of rapid variation in relative humidity induced by dew or because of higher rainfall during this period (Gauhl 1994). Rainfall occurred on 6 days over this period and represented a total of 109 mm.

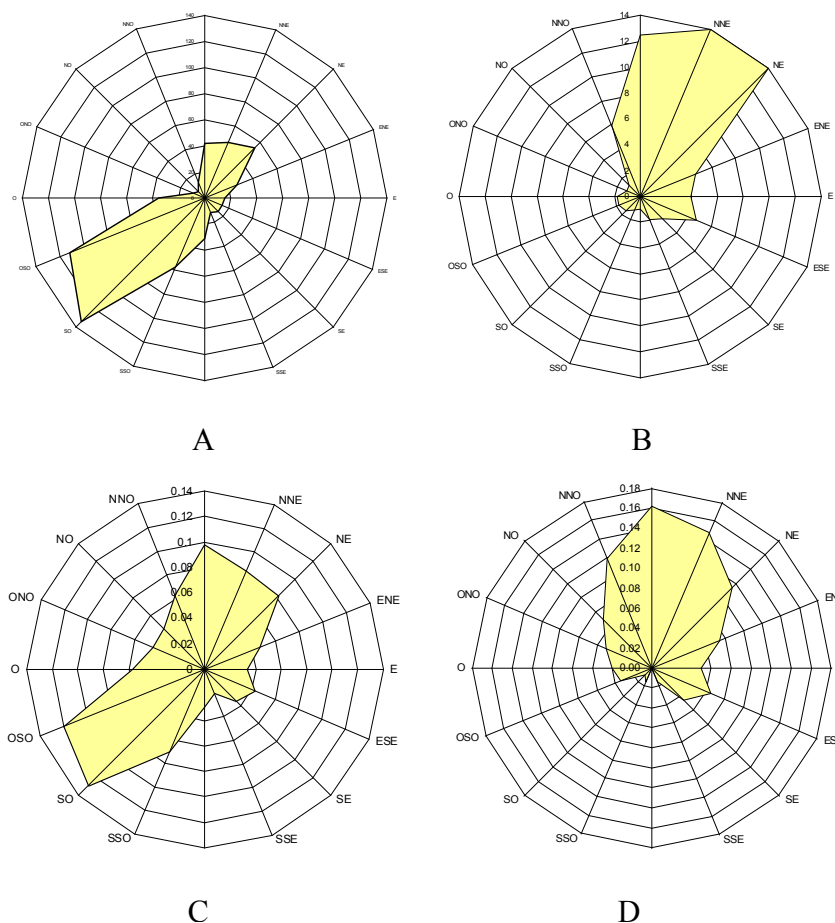


Figure 9.

The cumulative wind speed (the sum of wind speed over all measurements points at which the wind was blowing in that direction) and the frequency of wind record is given for each direction between D₀ (5/10/2010) and D₁₇ (19/10/2010) and was either calculated:

- Considering all daily data (A for the speed, C for the frequency)
- Considering only data recorded between 5 and 9 a.m (B for the speed, D for the frequency)

Axis	Mean distance travelled	Length of the axis (m)	Longest distance travelled (m)
A	273.129	900	800
B	104.583	1000	500
C	443.466	1000	1000
D	613.684	1000	1000
E	367.537	1000	900
F	133.473	1000	600
G	140.638	700	700
H	195.115	500	500

Axis	Mean distance travelled	Length of the axis (m)	Longest distance travelled (m)
a	2.566	25	5
b	5.797	25	10
c	3.4	25	7.5
d	0	25	0
e	0	25	0
f	2.6	25	5
g	3	25	5
h	7.647	25	12.5

Table 1. Mean and longest distance travelled along each of the axis of both ascospores (above) and conidia (bottom) experimental site. Mean distances were obtained from equation 7. Minimum and maximum values are highlighted in bold.

Statistical analysis

Ascospore dataset

A large variation in the mean distance travelled by spores was observed in the different directions of the experimental design, ranging from 104 m (axis B) to 613 m (axis D) (Table 1).

The geometric kernel provided a better fit than the exponential kernel (Table 2), probably because of its higher capacity to consider the LDD events detected (cf. Figure 10B and C). As a result, the estimated value of parameter b indicated that the shape of the distributions of ascospore dispersal distances is fat-tailed ($2 < b < 3$). The geometric model estimated the total number of ascospores released by the source at 2.26×10^7 . Estimation of both the directional density and mean distance functions (calculated using the geometrical kernel) revealed contrasting patterns of anisotropy in the dispersion of *M. fijiensis*. Indeed, anisotropy in density (Figure 11.A) appeared not to be significant (*i.e.*, σ not significantly different from 0, resulting in a uniform directional density function), whereas anisotropy in the mean distance travelled (Figure 11.B) was significant (*i.e.*, k significantly different from 0). This indicates that although there is no significant preferential direction taken by spores, they do disperse further in the SW direction (angle $\nu = 4.19$ radians). This result was inconsistent with the mean distances calculated from raw data (Table 1). The map of probability for banana plants to be infected (infectious potential) is shown either in 2-D or in 3-D in Figure 12. This representation corresponds to the IDF estimated from the experiment. Finally, the average dispersal distance value (over the various directions) calculated from this infectious potential distribution was 210.41 m. Note that this calculation takes into account the fact that spores are deposited over an increasing width sector. In other words, it explicitly includes the concept that, because the number of sampling directions is finite and constant, the probability of missing a dispersal event increases with distance from the source.

	Exponential	Geometric
K	6	7
LL	-836.12	-737.91
S_0	9.70E+05	2.26E+07
μ	5.46	4.79
σ	0.23	0.33
g_0	6.59	3.61
ν	4.42	4.19
k	0.60	1.38
b		2.0016
AIC	1684.23	1489.83

Table 2. Ascospores experimental site: Maximum likelihood estimated of the different parameters considering a negative binomial underlying model and either an exponential or a geometric kernel of dispersal. K for the number of parameters, LL for maximum likelihood, S_0 is the source strength, μ and ν (radian) are the direction of the density and distance anisotropy functions respectively, σ and k are the variability around the mean for the density and distance anisotropy functions respectively, g_0 is a constant that cannot be compared between different kernel and b is the parameter of the shape of the tail. AIC was calculated as $AIC = 2 \cdot K - 2LL$. Significant values of both anisotropy functions (tested using a LLRT) are highlighted in bold.

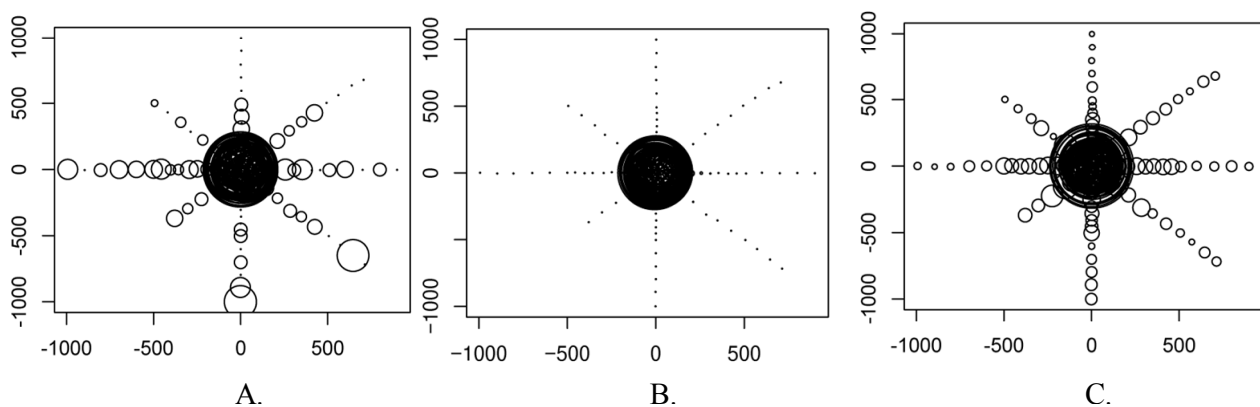


Figure 10. Ascospores experimental site: Comparison between the total density of resistant lesions observed (TDRL) (A) and predicted either by the exponential (B) or the geometric (C) kernel of dispersal (see table 2 for parameter values of both functions). TDRL is expressed in a “log (1+ of the value)” scale. Smallest dots represent a TDRL = 0.

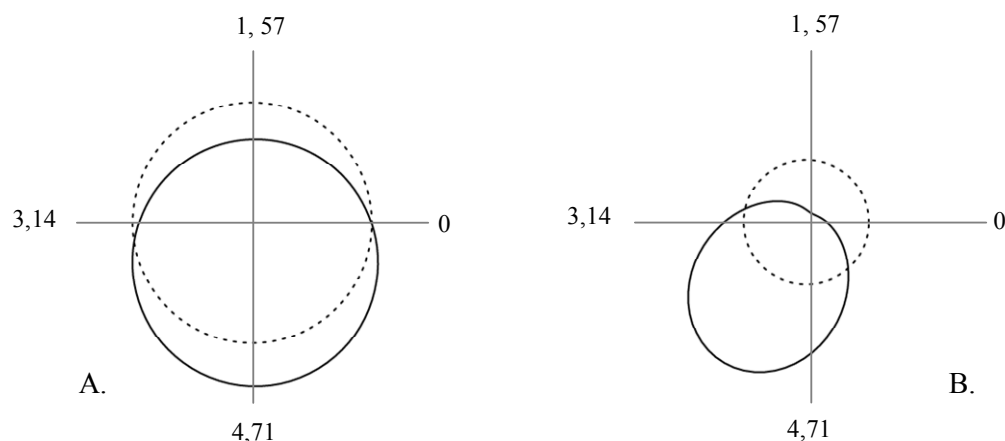


Figure 11. Ascospores experimental site: Estimates of the anisotropy functions. Both estimated directional functions (solid lines, A. for the density and B. for the mean distance function) can be compared to their equivalent uniform functions (dotted lines). Here anisotropy in density was not significant whereas anisotropy in mean distance was significant. Directional angles are given in radian.

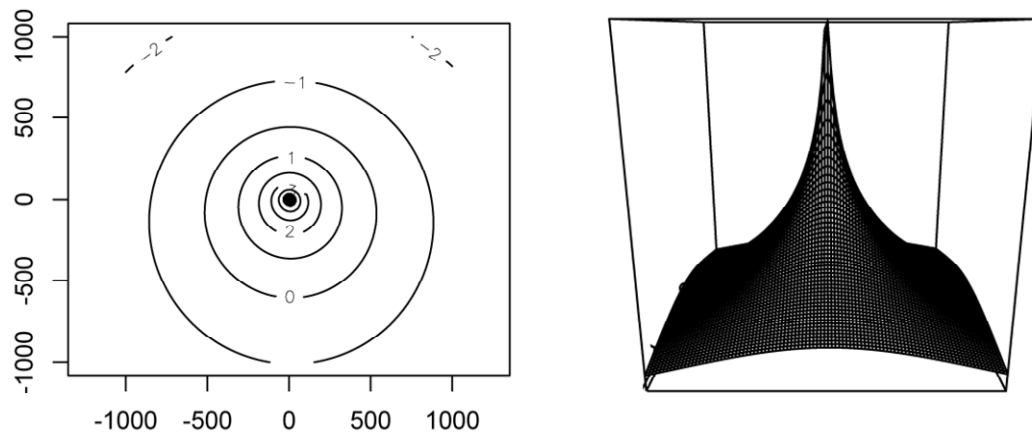


Figure 12. Ascospores experimental site: 2D and 3D probability of infection of banana plants maps.

Conidia dataset

The same analyses were applied to data from conidia. The mean distance travelled by spores varied between 0 m (axes D,E) and 7.6 m (axis H) (Table 1). Here, the exponential kernel provided a better fit than the geometric kernel (Table 3). The exponential model estimated the total number of conidia released by the source at 5×10^3 ascospores. Considering the exponential kernel, anisotropies in both density and distance displayed a significant signal in the SW direction (axis H). The estimated mean distance of dispersal for conidia calculated from the infectious potential function was 3.12 m.

	Exponential	Geometric
K	6	7
LL	-65.66061	-69.68823
S_0	5.64E+03	5.01E+03
μ	4.24	4.09
σ	1.99	1.37
g_0	4.92	13.49
ν	4.07	4.11
k	2.13	2.03
b		60080
AIC	143.32	153.38

Table 3. Conidia experimental site: Maximum likelihood estimated of the different parameters considering a negative binomial underlying model and either an exponential or a geometric kernel of dispersal. K for the number of parameters, LL for maximum likelihood, S_0 is the source strength, μ and ν (radian) are the direction of the density and distance anisotropy functions respectively, σ and k are the variability around the mean for the density and distance anisotropy functions respectively, g_0 is a constant that cannot be compared between different kernel and b is the parameter of the shape of the tail. AIC was calculated as $AIC = 2 \cdot K - 2 \cdot LL$. Significant values of both anisotropy functions (tested using a LLRT) are highlighted in bold.

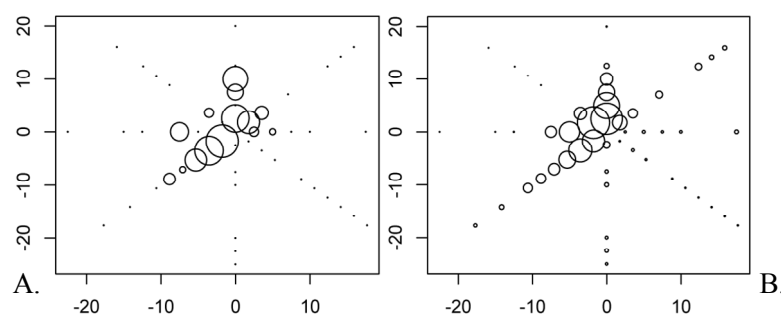


Figure 13. Conidia experimental site: Comparison between the total density of resistant lesions observed (TDRL) (A) and predicted by the exponential kernel of dispersal (B, see table2 for parameter values). TDRL is expressed in a “log (1+ the value)” scale. Missing dots (empty space) represents the missing data due to mortality and the smallest dots indicates plants without resistant lesions.

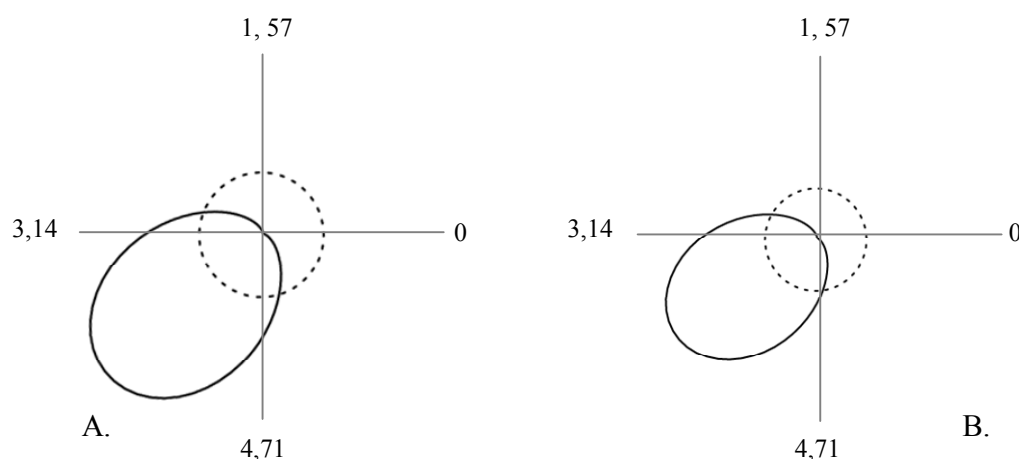


Figure 14. Conidia experimental site: Estimates of the anisotropy functions. Both estimated directional functions (solid lines, A. for the density and B. for the mean distance function) can be compared to their equivalent uniform functions (dotted lines). Both anisotropies in density and in distance were significant. Directional angles are given in radian.

Correlation between density, distance and wind patterns.

Correlation between spore dispersal (in terms of both density and distance) and wind patterns in each direction (in frequency or speed, and either for all recorded points or points recorded between 5 and 9 a.m.) provided contrasting results for ascospores and conidia. Indeed, whereas no significant correlation was found in the case of ascospores, we found a positive linear correlation ($R=0.91; P=0.0001$ & $R=0.95; P=0.00001$ between wind frequency in a given direction and density or mean distance travelled, respectively) (Table 4).

Ascospores	distance travelled	density
Wind frequency (total)	0,06 (NS)	0 (NS)
Wind speed (total)	0,36 (NS)	0,02 (NS)
Wind frequency (5-9 a.m)	0,50 (NS)	0,50 (NS)
Wind speed (5-9 a.m)	0,65 (NS)	0,57 (NS)
Conidia	distance travelled	density
Wind frequency (total)	0,91 (***)	0,95 (***)
Wind speed (total)	0,69 (NS)	0,70 (*)
Wind frequency (5-9 a.m)	0,29 (NS)	0,33 (NS)
Wind speed (5-9 a.m)	0,09 (NS)	0,28 (NS)

Table 4. Spearman rank correlation coefficient (and the associated two-sided p-value, NS >0.05 * $P<0.05$; *** $P<0.0001$) between spores density or distances travelled and wind patterns in a given direction. Significant values are highlighted in bold.

DISCUSSION

Originality of the study

Given their biological significance, measuring both the shape and the tail of the distribution of dispersal distances in natural populations of plant pathogens would seem to be essential (Kot et al. 1996, Austerlitz 1997, Nathan 2001, Excoffier et al. 2009). In similar vein, a growing quantity of data has been published recently in the study of pollen dispersal curves (Klein et al. 2006, Busi et al. 2008, Darmency et al. 2009). However, despite the apparently similarity with pollen in their dispersal mechanisms (*i.e.*, wind-borne propagules), very few studies have been successfully realised in the case of aerial fungal plant pathogens. In the present study, we combined two field plot experiments with an adapted statistical methodology of analysis of spore dispersal anisotropy to estimate, independently and successfully, the kernel of dispersal of both ascospores and conidia in the banana plant pathogen fungus *Mycosphaerella fijiensis*. The experimental design specially set up for this purpose differs from the majority of previous studies carried out on fungal plant pathogens in several aspects, particularly in two major points: i) specific precautions were taken to avoid contamination either from external spores or from secondary foci. Indeed, we used a marked source of inoculum composed of specific fungicide-resistant strains in an environment where those strains were not present, and we measured the subsequent disease gradient over a single-generation of the pathogen. ii) The study benefitted from large sample size and the originality of the sampling design, which was adapted to the detection of LDD events and of anisotropy. Indeed, in the ascospore experimental field, banana plants were installed for up to 1000 m and along 8 directions radiating from the inoculum source, with an intensive strategy of tight trap disposition up to 100 m that was subsequently relaxed while increasing the receptive surface with distance. In comparison, previous fungal plant disease gradient studies have generally been realized i) at a small scale (*i.e.*, distances of < 50 m from a source of inoculum), ii) without using marked inoculums, iii) during several life cycles of the pathogen, and iv) with decreased sampling effort with distance (Mundt 1989, Fontem et al. 1991, Inman et al. 1999, Cox & Scherm 2001 and references therein). Many of these studies were also limited in the number of sampling locations (most had less than 20 sites). As a result, no dispersal kernels were properly estimated from these studies. Indeed, the occurrence of contamination (from either external spores or secondary foci) could never be excluded, thus precluding the detection of rare LDD events. In the literature, we found only two studies that managed to measure a “large scale” single-generation disease gradient from an identified source of inoculum in a windborne pathogen fungus. In the first of these studies, Sackett and Mundt (2005) studied dispersal in *Puccinia striiformis* causing wheat stripe rust up to 170 m in one-dimension (up- and downwind directions) from a primary source of a specific strain. This experiment allowed these authors to detect rare LDD events 80 m from the source, to fit several dispersal kernels and to determine the non-exponentially bound nature of the gradient tails. However, the one-dimensional nature of the field plots did not allow the authors to test for anisotropy in dispersal patterns. The second reference study (Soubeyrand et al. 2007) was carried out on the same pathosystem but focused specifically on the detection of anisotropic patterns. For this purpose, the authors used a source comprising a specific strain and assessed a disease gradient up to 400 m through a large array of small plots arranged randomly in two-dimensions around the source plot. Combined with

the introduction of a new statistical framework, this study allowed the detection of rare LDD events at 225 m and the description of spore dispersal anisotropy in both density and distance. However, the weight of the tail shape was not characterized since a unique exponential kernel was fitted to these data. Here, we extended this type of statistical methodology to the geometrical kernel family in order to simultaneously estimate the shape of the dispersal distance distribution (including the weight of the tail) and test patterns of anisotropy in either density or distance in BLSD disease gradients.

Ascospore experimental site

In the ascospore experimental site, the trap layouts used allowed us to observe disease symptoms over the majority of the experimental area (87.5% of the sites) without any strong presupposition as to the average distance of dispersal. The success of this approach was probably also due to the high strength of the artificial source installed in the centre of the experimental design. Indeed, both the experimental results and the model estimations indicated the high number of ascospores potentially released (*i.e.*, 7 and 22 million, respectively). However, our experimental measurements provide a temporal estimate of the number of ascospores released whereas the model estimates the total number of ascospores released during the entire experimental period (taking into account several periods of ascospore release). To our knowledge, this is the first example of the use of such a massive artificial inoculum source in a plant pathogen system. In previous studies on *M. fijiensis* (Rutter et al. 1998, Abadie et al. in prep), a small number of diseased leaves was installed at different points inside a banana plot. Here, we managed to concentrate the potential inoculum of 50 banana plants into a tiny area. Moreover, because the source of inoculum was composed of a high frequency of strobilurin-resistant strains, and the banana trap plants were treated with azoxystrobin, external contamination was reduced significantly. However, external contamination of the plants was detected over longer distances in the experimental design (*i.e.*, far from the source) through intensive molecular assessment of the resistance. We surmised that these contaminations came from banana plantains located in the region surrounding the experimental site but since the experiment was not designed to quantify external contamination, no attempt was made to interpret their different spatial levels or to trace it back to its origin.

Analysing the spatial patterns of resistant lesions alone allowed us to detect both a steep gradient in the first few metres of each direction as well as rare LDD events up to 1000 m from the source in two of the eight directions. Such rapid and sharp gradients have been described previously in gradient studies of airborne fungi (Gregory 1968, Sacket and Mundt 2005) including *M. fijiensis* (Burt et al. 1997, Abadie et al. in prep). However, the detection of dispersal events up to 1 km from a localized primary source of inoculum far exceeds the distance over which LDD has been characterized previously in plant pathogen fungi (Sackett and Mundt 2005, Soubeyrand et al. 2007). However, comparisons between *M. fijiensis* and *Puccinia striiformis* are not very useful since both the natural properties of the spores and the underlying release mechanisms differ (Brown 1997). In the case of *M. fijiensis*, previous results obtained by Burt et al. (1998) and Amil et al. (2007) suggested that ascospores could be dispersed over several kilometres but the occurrence of

contamination (either external or due to several foci) could not be excluded. Here, our results clearly show that ascospores of this pathogen can be dispersed at least up to 1 km.

Statistical analysis allowed the simultaneous estimation of gradient curves (measurement of the decrease in intensity with distance from the source) and anisotropy functions (assessment of which dispersal directions are preferred by spores and how far, on average, a spore travels in various directions). It is important to note that the statistical analysis was carried out under the following assumptions: i) all spores were dispersed following the same kernel, and ii) the average leaf receptivity in a trap plot was the same over the entire experimental area.

Here, disease gradient was clearly a better fit to the geometric kernel (*i.e.* the non-exponentially bound model) than to the exponential model, and the estimated value of the shape parameter well described the fat-tailed nature of the distribution (slower decrease than in the exponential case). This result is consistent with many theoretical considerations showing that wind dispersal is a major mechanism of LDD, and that windborne fungi frequently display such a pattern (Gregory 1945, Brown 1997, Wingen 2007). It is interesting to note that the relation between the log of the total density of resistant lesions and distance appeared to be linear up to 150 m (see Figure 7b). Such a result is characteristic of an exponential decrease at this spatial scale. In other words, our conclusions would have been different if we had carried out our experiment up to 150 m only. This result indicates that the spatial scale considered in the present study was appropriately adapted to the dispersal capacity of *M. fijiensis*. However, as pointed out by Sackett and Mundt (2005), this observation also serves as a general illustration of how the relative suitability of dispersal models can be obscured when data are measured along a truncated part of the dispersal distance distribution.

Anisotropy functions indicated that there was no significant preferred direction taken by ascospores, although they dispersed on average further in the SW direction. However, it is important to keep in mind that the von Mises functions used for the analysis are unimodal, which means that they can characterize only one unique preferred direction. Indeed, the SW anisotropy in distance estimated from the statistical framework is probably a bias resulting from the higher mean distances travelled in both the South and West directions. Further analyses taking a multimodal function of anisotropy into account will be required to settle this issue. In their study, Soubeyrand et al. (2007) detected a significant signal for both anisotropies that did not coincide (different directions). These authors hypothesized that several factors, such as topography, wind speed, wind direction and other climatic conditions, may be implicated in this difference but further advances in understanding of this phenomenon were precluded by lack of such data. Here, given the observed wind patterns, and considering that the direction taken by an ascospore is fixed by the direction of the wind at the moment of expulsion, our results suggest that several dispersal events might have occurred at different hours of several days, *i.e.*, not only during the time interval of 5-9 a.m.—the period when ascospore discharge is thought to be at its peak (Gauhl 1994). Indeed, both the frequency and the cumulative speed of the wind during this time period were both highest in the North and North-East directions. Different wind directions at the exact moments of different ascospore discharge events could explain the absence of anisotropy in density. However, the higher distances travelled in the

South and West directions are probably due to the highest wind speeds being in those directions, but no definitive conclusions can be drawn regarding this point because we do not know the exact periods when ascospore discharge occurred.

Conidia experimental site

In the conidia experimental site, we observed symptoms in a minor fraction (31%) of the banana plant traps, and only up to 12.5 m from the source plot. We detected a difference in the expected number of conidia released by the source between the estimation obtained from the model (around 5000) and estimates from field measurements and laboratory counts (around 150,000). This result seems to indicate that the majority of conidia did not disperse out of the central source plot. Unfortunately, we do not have any measure of conidia that dispersed at the scale of the source. Rain splash may have limited the dispersion of conidia (Gauhl 1994) at the source scale (*i.e.*, at distances < 2.5 m). However, another plausible explanation could be that many conidia were not dispersed because they could not be dislodged due to insufficient wind forces. The fact that anisotropy in density and in distance are oriented in the same direction as the highest winds reinforces this suggestion. Lastly, conidia could also have been lost during transport from the location at which the source plants were produced and the experimental plot. We detected contamination by sensitive strains (30% of the total number of lesions counted). However, unlike the situation at the ascospore site, this contamination was not detected at the border of the experimental design. Moreover, since no BLSD lesion was detected further than 12.5 m from the source, these sensitive strains probably originated from the central source itself, which was composed of 96% resistant lesions according to molecular analysis. Heterogeneity in sporulation between resistant and susceptible strains or stochastic distribution within the small population of lesions counted (255 in total), could explain this observation.

Analysing spatial patterns of resistant lesions allowed us to detect a gradient in 1 of the 8 axes only (axis h), and only up to 12.5 m. In other directions, either no lesions were detected or lesions were found only on the first plants surrounding the source without displaying any gradient pattern from the centre. It is not possible to compare our results with previous data on this pathogen since the only studies on dispersal of conidia thus far have been based on air trapping methods within the banana canopy without using a specifically marked inoculum source (Rutter and Burt 1997). However, our results are consistent with those of many other studies, either in the field or under simulated conditions, showing that conidia are often dispersed through short distances gradient over just a few metres from the source (Rapilly 1991, Willocquet et al. 1998, Cox & Scherm 2001, Grove 2006).

Contrary to the ascospore analysis, disease gradient was better fitted by the exponential kernel than by the geometrical model. Further analyses considering thin-tailed kernels (*e.g.* Gaussian function) will be needed to determine whether the tail of the distribution decreases more quickly than with the exponential function. The detected anisotropy in distance and in density in the South direction was consistent with the mapped results and can be explained by the recorded wind patterns. Indeed, our

results suggest that, the higher the average wind speed in a given direction, the higher the number of conidia dislodged and the higher the distance travelled by conidia.

Comparison between ascospore and conidia dispersal

Several conclusions related to the similarities and differences between ascospores and conidia dispersal can be drawn from the present results. Firstly, our study indicates that both ascospores and conidia can be wind dispersed but that a contrasting kernel of dispersal can also be generated. Indeed, dispersal of ascospores displayed a fat-tailed distribution, whereas an exponentially bounded distribution appeared best adapted to describe dispersal of conidia. Regarding this point, further adjustments taking other theoretical dispersal kernels into account need to be performed. Secondly, even if the inferred average dispersal distances of 210 m for ascospores and 3 m for conidia cannot be compared directly because of different source intensities, our results confirm the suspicion often previously enounced that, in *M. fijiensis*, conidia are responsible for short-distance dispersal whereas ascospores are responsible for spread of the disease over longer distances (Gauhl 1994, Rutter et al. 1998). Differences in both physical properties and air release mechanisms are probably at the origin of the contrasting patterns of dispersal of ascospores and conidia (Rapilly 1991). Finally, our results indicate that contrasting patterns of anisotropy (in both density and distance) between ascospores and conidia can be obtained in the same environment depending on the prevailing winds at the time of ascospore discharge. In this specific context, further analyses considering a multimodal function of anisotropy should be carried out.

CONCLUSION

In this study, we used a direct method to study the natural gradients of dispersal of *M. fijiensis*. Such demographic approaches have previously been criticized for two main reasons. The first is that direct methods generally inform on the dispersal processes of species but not on gene flow. Here, we addressed this drawback by analysing disease spread, which is a measure of “effective dispersal” of the pathogen because it includes both physical transport and establishment at the new site. The second criticism is that direct measurements are constrained in time and space, since they can be affected by the time frame during which observations are made and by difficulties in detecting long-distance movement. Here we showed that it is possible to characterize LDD by using a specially adapted experimental design. However, the measured dispersal characteristics are still thought only to be valid for the specific environmental conditions of the area and the time period of this experiment. In this context, it would be interesting to repeat this study in another location to disentangle the relative influence of environmental conditions and biological characteristics on patterns of *M. fijiensis* dispersal and at a distance > 1Km from the source. Finally, the two independent kernels inferred to describe ascospores and conidia dispersal in the present study should help parameterize future theoretical work aimed at predicting both the speed of the wave of BLSD disease expansion in new areas and spatial patterns of evolution of fungicide resistance under different management strategies.

PROSPECTS

The results presented and discussed here are based on preliminary analyses. Several improvements should be integrated in the near future. First, new theoretical kernels and multimodal anisotropy functions should be included in the analysis. Second, it will be interesting to fit kernels on the different axis independently to test whether or not the nature of the kernel depends on different wind patterns. Finally, the present analyses assumed an underlying negative binomial distribution to model the number of lesions on each leaf. Such a model is known to be a robust descriptor of dispersal events observed at different distances from a source of inoculum. However, in the present study we did not count specifically all lesions resulting from the source but rather estimated this number from the percentage of resistant genotypes over a maximum of 32 lesions. This estimation should be robust in most of the sites but it is possible that in some situations (specifically where a high number of lesions was counted and a small percentage of resistance were detected at some long distance sites), the estimated number of resistant lesions might have been biased. Further analyses considering an underlying model taking into account such a bias in the estimation of resistant individuals (*e.g.*, an hypergeometric law) will be required.

Acknowledgments

We thank SOCAPALM for provide us with the opportunity to implant our experimental design on their parcels, Rémy Habas for the design and construction of the artificial banana plants used as the source of ascospore dispersal, Bayer Crop Science for assistance with the molecular assessment of resistance, and CARBAP station staff members who where instrumental to the success of this study.

REFERENCES

- Abadie C, Bonnot F, Lannou C, *et al.* (*in prep*) Measurement of the ascospores dispersal of the fungal pathogen of bananas *Mycosphaerella fijiensis* at a local scale.
- Amil AF, Heaney SP, Stanger C, Shaw MW (2007) Dynamics of QoI sensitivity in *Mycosphaerella fijiensis* in Costa Rica during 2000 to 2003. *Phytopathology* **97**, 1451-1457.
- Austerlitz F, Dick CW, Dutech C, *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* **13**, 937-954.
- Austerlitz F, JungMuller B, Godelle B, Gouyon PH (1997) Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology* **51**, 148-164.
- Bartlett D, Clough JM, Godwin J, *et al.* (2002) The strobilurin fungicides. *Pest Management Science* **58**, 649-662.
- Broquet T, Petit EJ (2009) Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology Evolution and Systematics* **40**, 193-216.
- Brown JKM (1997) *Plant pathogens and plant diseases* Australasian Plant Pathology Society Inc.
- Brown JKM, Hovmoller MS (2002) Epidemiology - Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537-541.
- Bullock JM, Shea K, Skarpaas O (2006) Measuring plant dispersal: an introduction to field methods and experimental design. *Plant Ecology* **186**, 217-234.
- Burie JB, Langlais M, Calonnec A (2011) Switching from a mechanistic model to a continuous model to study at different scales the effect of vine growth on the dynamic of a powdery mildew epidemic *Annals of Botany* **107**, 885-895.
- Burt PJA (2003) Airborne dispersal of *Mycosphaerella fijiensis*. In: *Mycosphaerella Leaf Spot Diseases of Bananas: Present Status and Outlook. Proceedings of the Workshop on Mycosphaerella Leaf Spot Diseases, San José, Costa Rica.*
- Burt PJA, Rutter J, Gonzales H (1997) Short-distance wind dispersal of the fungal pathogens causing Sigatoka diseases in banana and plantains. *Plant Pathology* **46**, 451-458.
- Busi R, Yu Q, Barrett-Lennard R, Powles S (2008) Long distance pollen-mediated flow of herbicide resistance genes in *Lolium rigidum*. *Theoretical and Applied Genetics* **117**, 1281-1290.
- Carlier J (2000) Population genetic structure and dispersal of *Mycosphaerella fijiensis*. *Info musa* **13**, 17-20.
- Churchill ACL (2010) *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Molecular plant pathology* **12**, 307-328.
- Cox KD, Scherm H (2001) Gradients of primary and secondary infection by *Monilinia vaccinii-corymbosi* from point sources of ascospores and conidia. *Plant Disease* **85**, 955-959.
- Darmency H, Klein EK, De Garanbe TG, *et al.* (2009) Pollen dispersal in sugar beet production fields. *Theoretical and Applied Genetics* **118**, 1083-1092.
- De Lapeyre de Bellaire L., Fouré E., Abadie C., J. C (2010) Black leaf streak disease is challenging the banana industry. *Fruits* **65**, 327-342.
- Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology Evolution and Systematics* **40**, 481-501.

- Fisher NI (1995) *Statistical Analysis of Circular Data*.
- Fontem DA, Berger RD, Weingartner DP, Bartz JA (1991) Progress and spread of dark leaf-spot in cabbage. *Plant Disease* **75**, 269-274.
- Forster GA, Gilligan CA (2007) Optimizing the control of disease infestations at the landscape scale. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 4984-4989.
- Fouré E (1987) Varietal reactions of bananas and plantains to black leaf streak disease. In: *Banana and Plantain Breeding Strategies, Proceedings of an International Workshop, Cairns, Australia, 13-17 October 1986* (ed. Persley GJaDL, E.A., eds).
- Gauhl F (1994) Epidemiology and ecology of black sigatoka (PhD Thesis), p. 120. INIBAP, Montpellier.
- Gilligan CA (2008) Sustainable agriculture and plant diseases: an epidemiological perspective. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 741-759.
- Gregory PH (1968) Interpreting plant disease dispersal gradients. *Annual Review of Phytopathology* **6**, 189-212.
- Grove GG, Biggs AR (2006) Production and dispersal of conidia of *Leucostoma cinctum* in peach and cherry orchards under irrigation in eastern Washington. *Plant Disease* **90**, 587-591.
- Inman AJ, Fitt BDL, Todd AD, Evans RL (1999) Ascospores as primary inoculum for epidemics of white leaf spot (*Mycosphaerella capsellae*) in winter oilseed rape in the UK. *Plant Pathology* **48**, 308-319.
- Jones DR (2000) Diseases of banana, abaca and enset. CABI Publishing, Oxon.
- Klein EK, Lavigne C, Picault H, Renard M, Gouyon PH (2006) Pollen dispersal of oilseed rape: estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology* **43**, 141-151.
- Kot M, Lewis MA, vandenDriessche P (1996) Dispersal data and the spread of invading organisms. *Ecology* **77**, 2027-2042.
- Lacey J (1996) Spore dispersal - Its role in ecology and disease: The British contribution to fungal aerobiology. *Mycological Research* **100**, 641-660.
- Lavigne C, Klein EK, Vallee P, *et al.* (1998) A pollen-dispersal experiment with transgenic oilseed rape. Estimation of the average pollen dispersal of an individual plant within a field. *Theoretical and Applied Genetics* **96**, 886-896.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**, 183-189.
- Lenormand T, Raymond M (1998) Resistance management: the stable zone strategy. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 1985-1990.
- Mundt CC (1989) Use of the modified Gregory model to describe primary disease gradients of wheat leaf rust produced from area sources of inoculum. *Phytopathology* **79**.
- Nathan R (2001) The challenges of studying dispersal. *Trends in Ecology & Evolution* **16**, 481-483.
- Rapilly F (1991) *L'épidémiologie en pathologie végétale. Mycoses aériennes*
- Robert S, Rieux A, Argout X, Carlier J, Zapater MF (2010) Optimized genotyping with microsatellite markers in the fungal banana pathogen *Mycosphaerella Fijiensis*. *American Journal of Botany* **97**, E130-E132.

- Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics* **173**, 1033-1045.
- Rohatgi VK (2003) *Statistical Inference*.
- Rutter J, Burt PJA, Ramirez F (1998) Movement of *Mycosphaerella fijiensis* spores and Sigatoka disease development on plantain close to an inoculum source. *Aerobiologica* **14**, 201-208.
- Saccheri IJ, Rousset F, Watts PC, Brakefield PM, Cook LM (2008) Selection and gene flow on a diminishing cline of melanic peppered moths. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16212-16217.
- Sackett KE, Mundt CC (2005) The effects of dispersal gradient and pathogen life cycle components on epidemic velocity in computer simulations. *Phytopathology* **95**, 992-1000.
- Scherm H (1996) On the velocity of epidemic waves in model plant disease epidemics. *Ecological Modelling* **87**, 217-222.
- Shaw MW (1995) Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. *Proceedings of the Royal Society of London Series B-Biological Sciences* **259**, 243-248.
- Sierotzki H, Parisi S, Steinfeld U, *et al.* (2000) Mode of resistance to respiration inhibitors at the cytochrome bc(1) enzyme complex of *Mycosphaerella fijiensis* field isolates. *Pest Management Science* **56**, 833-841.
- Skarpaas O, Shea K, Bullock JM (2005) Optimizing dispersal study design by Monte Carlo simulation. *Journal of Applied Ecology* **42**, 731-739.
- Skellam JG (1951) Random dispersal in theoretical populations *Biometrika* **38**, 196-218.
- Soubeyrand S, Enjalbert J, Sanchez A, Sache I (2007) Anisotropy, in density and in distance, of the dispersal of yellow rust of wheat: Experiments in large field plots and estimation. *Phytopathology* **97**, 1315-1324.
- Soubeyrand S, Held L, Hohle M, Sache I (2008) Modelling the spread in space and time of an airborne plant disease. *Journal of the Royal Statistical Society Series C-Applied Statistics* **57**, 253-272.
- Tufto J, Engen S, Hindar K (1997) Stochastic dispersal processes in plant populations. *Theoretical Population Biology* **52**, 16-26.
- Wilocquet L, Berud F, Raoux L, Clerjeau M (1998) Effects of wind, relative humidity, leaf movement and colony age on dispersal of conidia of *Uncinula necator*, causal agent of grape powdery mildew. *Plant Pathology* **47**, 234-242.
- Wingen LU, Brown JKM, Shaw MW (2007) The population genetic structure of clonal organisms generated by exponentially bounded and fat-tailed dispersal. *Genetics* **177**, 435-448.

DISCUSSION GÉNÉRALE

Discussion générale et perspectives

Cette thèse avait pour objectif l'étude des processus de dispersion et la quantification des flux de gènes qui en découlent chez le champignon pathogène des bananiers *Mycosphaerella fijiensis* à l'échelle locale d'un bassin de production Camerounais. Pour cela, une stratégie combinant différentes approches a été adoptée. Les différents résultats obtenus ont indépendamment été discutés et comparés aux résultats de la littérature dans les articles au sein desquels ils ont été présentés. Le but de ce chapitre est autre, il aspire plutôt à mettre en relation ces différents résultats afin de discuter de leur cohérence, de leurs implications ainsi qu'à présenter quelques pistes de recherches.

1- De grandes populations en déséquilibre génétique

Une première partie de ce travail de thèse a été consacrée à l'étude de la structuration génétique spatiale des populations de *M. fijiensis* sur la zone de production du Mounjo par une approche de génétique du paysage (Manel *et al.* 2003). L'objectif était de parvenir à déterminer les facteurs écologiques et environnementaux pouvant influencer les flux géniques chez ce champignon.

Pas d'influence des flux géniques contemporains sur la structure génétique spatiale des populations.

Sur la base de critères génétiques, nous avons mis en évidence la présence de deux entités distinctes au sein de la large zone étudiée, l'une d'entre elles étant située à l'extrême Nord-Ouest de la zone d'étude (Chapitre 1). Cependant, aucun effet du paysage sur la SGS n'a pu être détecté, que ce soit pour des barrières potentielles aux flux géniques de type physique (montagne, rivière, route, zones urbaines), chimique (plantations de bananes traitées aux fongicides) ou biologique (zones dans lesquelles la densité de l'hôte est faible), puisque tous ces éléments apparaissent être intégrés au sein de la même unité génétiquement homogène. De la même façon, aucun effet de la distance géographique sur la SGS n'a été décelé. Un tel constat apparaît étonnant. En effet, bien que nos résultats indiquent que *M. fijiensis* présente d'importantes capacités naturelles de dispersion (e.g., nous avons observé des migrants de première génération jusqu'à 1Km de la source d'inoculum dans l'expérimentation directe présentée dans le Chapitre 3), ces distances demeurent inférieures à la surface totale occupée par les populations et ne devraient pas permettre un brassage homogène des fréquences alléliques sur l'ensemble de la zone.

Nous pensons que les processus d'expansion historique du pathogène au sein de la zone ainsi que des effectifs actuels importants sont à l'origine de l'absence de structuration génétique observée sur la plus grande partie du domaine d'étude. En effet, les résultats obtenus grâce à la méthode MSVAR adaptée au scénario « fondation-explosion » confirment que les populations de *M. fijiensis* ont subi une perte drastique de diversité génétique lors de l'introduction historique du pathogène en Afrique (voir Annexe 1). Par la suite, comme cela a été théoriquement démontré (Maruyama & Fuerst 1985,

Boileau *et al.* 1992), la croissance rapide des populations et l'existence de grandes tailles de populations (tous deux suspectées chez *M. fijiensis*) expliquent probablement que la perte de diversité engendrée puisse mettre énormément de temps à s'estomper. Les comptages de spores réalisés lors de l'expérimentation directe confirment cette hypothèse. En effet, nous avons observé que lors d'un évènement de libération, le nombre d'ascospores relarguées dans l'air peut être de l'ordre de $5.10^5/\text{m}^2$ de nécrose (ce qui correspond grossièrement à $\sim 2.5.10^5/\text{bananier non traité}$). Il est certain que tous ces individus ne participent pas à la génération suivante, mais cette observation indique tout de même l'existence d'un nombre important d'individus reproducteurs à l'échelle de la zone d'étude, nombre largement supérieur à l'effectif efficace calculé de façon théorique (voir Annexe 1). Dans une telle situation, l'effet de la dérive génétique devient négligeable et un grand nombre de générations (d'autant plus important que le nombre d'individus reproducteurs actuel est grand) est nécessaire afin de pouvoir accumuler de nouvelles mutations et ainsi ré-atteindre l'équilibre génétique. De cette façon, on peut s'attendre à ce que l'effet des barrières aux flux géniques ou bien de la distance géographique mette également énormément de temps avant de devenir détectable par les approches génétiques (Balkenhol *et al.* 2009, Biek & Real 2010). En d'autres termes, il est probable que certains éléments du paysage présents au sein de la zone influencent réellement les flux géniques mais les approches de génétique des populations ne nous permettent pas de le détecter, tout du moins avec la puissance statistique générée par nos marqueurs microsatellites.

Une étude théorique a été conduite récemment afin de déterminer quel était le temps nécessaire pour qu'une nouvelle barrière aux flux géniques devienne détectable par plusieurs approches de génétique spatiale des populations (Landguth *et al.* 2010). Les résultats indiquent que d'une manière générale, l'effet d'une nouvelle barrière est rapide (1-15 générations). Cependant, les populations simulées dans cette étude étaient initialement à l'équilibre mutation-dérive (en présence de la barrière) et l'effet des tailles de populations n'a pas été pris en compte. Il serait donc intéressant de réaliser un tel travail théorique dans le cas particulier de populations non à l'équilibre génétique (en simulant un scénario de fondation-explosion par exemple) et en faisant varier la taille de ces populations. Un tel exercice permettrait de préciser les conditions pour lesquelles les flux géniques contemporains peuvent réellement être caractérisés par l'étude de la diversité génétique neutre chez des organismes comme *M. fijiensis*.

Une discontinuité génétique historique qui s'estompe progressivement

Nous avons expliqué l'origine de la discontinuité génétique détectée comme étant la conséquence de la colonisation historique de *M. fijiensis* au sein de la zone. Selon cette hypothèse, l'effet des flux de gènes historiques sur la SGS pourrait persister pendant de nombreuses générations (plus de 200 dans le cas présent) dans des systèmes non à l'équilibre et caractérisés par des tailles efficaces importantes. Nous avons suggéré que la discontinuité génétique pouvait s'expliquer par l'occurrence de deux scénarios contrastés : le contact secondaire entre deux populations en

expansion (une venant du Sud et une venant du Nord) vs. un phénomène de surf génétique le long d'un axe unique d'expansion. Cependant, nous n'avons pas pu distinguer ces deux événements sur la base des données génétiques dont nous disposons. Tout d'abord, la grande variance dans les valeurs de F_{ST} observée est attendue dans les deux scénarios (Excoffier *et al.* 2009). Deuxièmement, l'étendue spatiale de la population du Nord est trop restreinte pour pouvoir analyser la structuration spatiale de la diversité génétique et déterminer si ce patron s'avère être caractéristique d'un scénario plutôt qu'un autre (cf. Figure 12). Troisièmement, il n'existe aucune prédiction théorique sur les caractéristiques de patrons de clines qui pourraient être mis en place lors d'un événement de surf génétique et qui auraient éventuellement pu permettre la distinction entre les deux scénarios par l'approche basée sur les clines neutres (grâce à la position relative des centres de clines observés à plusieurs locus par exemple, cf. article 3). Finalement, il a récemment été montré par simulations que la probabilité d'occurrence du surf génétique i) diminuait drastiquement lorsque la dispersion était modélisée par un noyau caractérisé par une queue lourde mais ii) augmentait dans le cas d'un rétrécissement spatial de la zone d'expansion (i.e., corridor) (Fayard *et al.* 2009). Ces résultats nous amènent à fournir des prédictions contradictoires quand à la nature du scénario à l'origine de la discontinuité génétique détectée. En effet, nous avons montré lors de l'expérimentation directe que la dispersion des ascospores était caractérisée par une fonction à queue lourde (Chapitre 3) mais également que la discontinuité génétique détectée semble justement être située dans une vallée représentant un tel corridor (Halkett *et al.* 2010) (cf. Figure 1 de l'article 3 et Figure 7 de l'introduction).

Il me semble important de parvenir à distinguer ces deux scénarios. En effet, bien qu'ayant été caractérisé sur des bactéries en boîtes de pétri (Hallatschek *et al.* 2007), le phénomène de surf génétique n'a jusqu'alors jamais été observé en conditions naturelles. Arriver à prouver son occurrence au sein de la zone étudiée au cours de cette thèse serait donc un résultat original. Dans un tel contexte, deux perspectives d'études m'apparaissent prometteuses. La première concerne un échantillonnage récemment réalisé au sein de la zone de fracture génétique pour lequel nous avons élargi la zone de prélèvement à 30 km au Nord et au Sud de la cassure génétique et augmenté le nombre d'individus prélevés par sites (30 au lieu de 3 ou 8). Le génotypage (en cours) et l'analyse future de cet échantillon pourront éventuellement nous permettre de différencier entre les deux patrons de diversité génétique présentés dans la Figure 12. La seconde fait référence à la formalisation théorique des conséquences du surf sur la diversité génétique. Beaucoup de travaux ont récemment été réalisés dans ce domaine (voir Excoffier *et al.* 2009 pour revue, ou Fayard *et al.* 2009) mais aucune étude ne s'est focalisée sur les patrons de mise en place de clines multilocus à des marqueurs neutres suite à un événement de surf génétique. Il serait d'ailleurs intéressant de comparer de tels patrons de clines à ceux attendus suite à un contact secondaire afin de tester l'existence de potentielles différences.

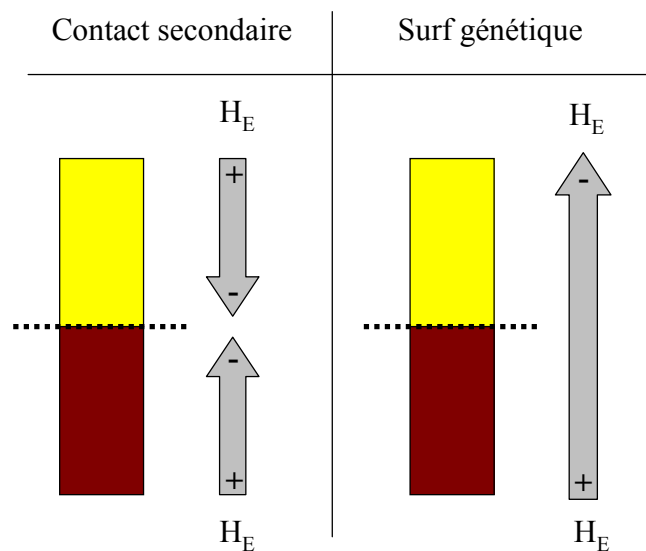


Figure 12. Illustration schématique de la distribution spatiale de la diversité génétique (H_E) théoriquement attendue suite à l'occurrence d'un contact secondaire entre populations ou d'un événement de surf génétique le long d'un axe d'expansion.

2- Des capacités de dispersion non négligeables

Parmi les différentes approches considérées durant cette thèse, deux d'entre elles nous ont permis de quantifier l'intensité des flux de gènes contemporains chez *M. fijiensis*. La première estimation, de type indirecte, est basée sur l'analyse de la variation temporelle de la largeur de clines observés à 15 marqueurs microsatellites au niveau de la zone de discontinuité génétique. Elle indique une distance moyenne de dispersion $\sigma = 1175$ mètres/génération sexuée avec un intervalle de confiance de [636m-1660m] (cf. article 3). La seconde s'appuie sur la mesure directe de deux distributions des distances de dispersion à partir d'une source d'inoculum primaire indépendante pour les ascospores et les conidies. Pour les conidies, nous avons observé une distance moyenne de dispersion $\delta = 3,12$ mètres/génération asexuée avec un minimum de 0 et un maximum de 7,6 m selon l'axe considéré. Pour les ascospores, la distance moyenne obtenue est de 283 mètres/génération sexuée avec un minimum de 104 m et un maximum de 613m selon la direction (cf. article 5).

Des estimations cohérentes

La distance moyenne de dispersion obtenue de façon indirecte ne permet pas de distinguer entre la dispersion des spores sexuées et asexuées. En effet, le génotypage des individus a été réalisé à partir de jeunes lésions pouvant aussi bien être issues de la germination d'une ascospore que d'une conidie. Cependant, pour l'inférence, le coefficient de diffusion obtenu combine donc les effets cumulés des dispersions des ascospores et des conidies. Il est exprimé en mètre par unité de temps et l'unité de temps que nous avons choisie est celle de la génération sexuée, c'est-à-dire le temps moyen entre le dépôt d'une ascospore sur une feuille et la production de nouvelles ascospores suite

à l'infection des tissus foliaires. Nous discuterons de l'impact d'un tel choix mais comparons dans un premier temps l'estimation indirecte vs. directe des ascospores.

Tout d'abord, les valeurs estimées par les deux approches apparaissent être du même ordre de grandeur (~ 4 fois supérieur dans le cas de l'estimation indirecte). D'une manière générale, très peu d'études ont explicitement comparé des estimations de la dispersion obtenues chez différents organismes à l'aide de méthodes génétiques et démographiques. Dans ce contexte, certaines relatent des estimations similaires (facteur de différence entre ~ 3 en faveur de la méthode génétique) (Oddou-Muratorio & Klein 2008, Sumner *et al.* 2001, Vandewoestijne & Baguette 2004) alors que d'autres décrivent des différences un peu plus importantes (facteur ~ 10 toujours en faveur de la méthode génétique) (Mallet *et al.* 1990, Barton & Gale 1993). Au vu des nombreuses différences sous-jacentes aux deux approches considérées (Tableau 3), nos résultats (facteur 4 observé en faveur de l'estimation indirecte) sont cohérents avec ceux précédemment publiés. Cependant, il me paraît important de discuter des différences attendues entre les deux approches dans le but de mieux comprendre l'origine de la variation observée.

	Clines neutres	Estimation du noyau de dispersion	
Type d'approche	Indirecte	Directe	
Pré-requis	2 entités génétiques spatialement contigues (issues d'un contact 2nd ou surfing génétique)	1 source marquée de dispersion + pièges disposés en 2D	
Nombre de générations	15	1	
<i>a priori</i> sur le noyau	Oui (Gaussien)	Non	
Ascospores & Conidies	Non dissociées	Mesure indépendante	
Dispersion mesurée en	1 D	2 D	
Résultats obtenus	Ascospores + conidies	Conidies	Ascospores
Valeur mesurée	$\sigma = 1175\text{m/G}$	$\delta = 3\text{m/G}$	$\delta = 280\text{m/G}$
Intervalle de confiance *	[636 - 1660]	[0 - 7.8]	[104-613]
Caractérisation des LDD	Non	Oui - $d_{\text{max}}=12.5\text{m}$ Noyau de type exponentiel	Oui - $d_{\text{max}}=1000\text{m}$ Noyau à queue lourde

* Pour la mesure directe, l'intervalle de confiance donné n'a pas été estimé de façon statistique. Il correspond simplement à la valeur minimale et maximale observée sur les 8 répétitions (8 directions).

Tableau 3. Tableau synthétique illustrant certaines différences notables entre l'approche indirecte et directe considérée pour l'estimation des flux géniques durant cette thèse et résumant les valeurs de paramètres estimés.

La première source de différence concerne la fenêtre de temps sur laquelle l'estimation a été réalisée. En effet, lors de l'expérimentation directe, nous avons observé la distribution des distances de dispersion sur une seule génération alors que pour l'inférence indirecte nous avons estimé des flux géniques moyennés sur plusieurs générations (en considérant une dispersion constante dans le temps). Or, la détermination du nombre de générations séparant les deux échantillonnages n'est pas triviale. En effet, *M. fijiensis* exhibe une reproduction continue dans le temps (générations chevauchantes) et disperse via la dissémination d'ascospores et de conidies dont le temps de génération n'est pas identique. Sachant cela, le choix du nombre de 15 générations (basé sur le temps de reproduction des ascospores uniquement) séparant nos deux échantillonnages peut s'avérer discutable. En effet, les conidies jouent probablement un rôle dans l'évolution spatiale des fréquences alléliques même si, comme nous l'avons montré lors de la mesure indépendante d'un noyau de dispersion, leurs capacités de dispersion dans l'espace semblent être beaucoup moins

importantes que pour celles des ascospores. Si l'on considère que les conidies ont participé à l'aplatissement temporel du cline, alors le nombre de générations réel séparant les deux échantillonnages peut être plus important que celui considéré, et cela d'un facteur 2 dans le cas extrême, certes non réaliste, où seules les conidies disperseraient. Dans une telle situation, l'estimation indirecte de la distance moyenne de dispersion pourrait être diminuée d'un facteur $\sqrt{2}$ (~ 1.41).

La deuxième source de variation provient de l'influence des événements de dispersion longue distance (LDD). En effet, les résultats obtenus lors de l'expérimentation directe indiquent que le noyau de dispersion des ascospores est à queue lourde, c'est-à-dire qu'il reflète une quantité non négligeable d'événements de LDD. Il est de ce fait envisageable que l'estimation indirecte puisse être biaisée puisque dans cette dernière la dispersion a été modélisée en faisant l'approximation de diffusion, c'est-à-dire sans considérer l'occurrence d'événements de dispersion à longues distances. Ainsi, il est possible que la valeur de σ puisse être surestimée par l'approche clines neutres. En effet, dans le cas de clines à l'équilibre migration-sélection, des simulations ont montré que l'estimation pouvait être surestimée de ~ 2 fois dans le cas de distributions très leptokurtiques (Lenormand 1998, Rousset 2001). Il serait donc intéressant de pouvoir réaliser ce genre de simulations pour des clines neutres. Une telle procédure a été entreprise lors de cette thèse mais n'a malheureusement pas abouti. Le principe est de comparer des clines obtenus par simulations selon différentes distributions de dispersion ayant la même variance mais des kurtosis différentes. Dans un futur proche, j'espère pouvoir parvenir à réaliser ce genre de simulations. D'une manière générale, il sera alors possible de déterminer l'enveloppe des valeurs de kurtosis induisant un biais supérieur à un certain seuil dans l'estimation de σ par l'approche indirecte que nous avons développé. Plus spécifiquement, en considérant des valeurs de kurtosis calculées à partir de la distribution des distances de dispersion issue de l'expérimentation directe (article 5), ce résultat pourrait nous permettre de quantifier le biais réel associé à l'estimation indirecte des flux géniques chez *M. fijiensis* réalisée durant cette thèse.

La troisième source de variation apparaît être une conséquence des deux premières. En effet, puisque l'estimation de la distribution des distances de dispersion a été obtenue sur une seule génération, il est théoriquement nécessaire de reproduire 15 fois une telle distribution pour pouvoir la comparer d'une façon la plus juste possible avec l'estimation indirecte. En mathématique, la reproduction spatio-temporelle d'un noyau de dispersion porte un nom : il s'agit de faire une convolution. Une telle procédure est classiquement utilisée pour prédire les vitesses de propagation des espèces envahissantes lors de la colonisation d'un nouvel habitat (Scherm 1996). Dans un tel contexte, il a été montré que pour des noyaux de dispersion à queues lourdes, les vitesses de colonisations prédites étaient plus rapides que dans le cas d'un noyau de type exponentiel (Kot *et al.* 1996). Dans le cadre de la mesure directe réalisée dans cette thèse, ce résultat indique que la distance moyenne prédite sur 15 générations devrait être plus importante que celle calculée sur une seule génération comme observé chez *M. fijiensis*. En perspective, il serait donc intéressant de réaliser une telle convolution afin de quantifier l'intensité de la différence à laquelle on pourrait s'attendre.

En conclusion, bien que nous n'ayons pas pu calculer l'influence quantitative de ces trois biais potentiels sur les deux estimations, leur analyse qualitative semble suggérer une surestimation du paramètre σ obtenu de façon indirecte et prédit une plus grande valeur du paramètre δ obtenu de façon directe dans le cas des ascospores. En conséquence, il semble que le facteur de ~ 4 détecté entre les deux estimations ne puisse que diminuer ce qui renforce la cohérence entre les deux mesures indépendantes des flux géniques.

*Les apports de cette thèse sur la dispersion de *M. fijiensis**

L'expérimentation directe, basée sur le suivi de la dispersion d'individus génétiquement marqués dans l'espace durant une seule génération nous a permis de mesurer la distribution des distances de dispersion des spores sexuées et asexuées de *M. fijiensis* (Chapitre 3). Les résultats obtenus confortent l'hypothèse précédemment émise que les ascospores participent à une dispersion à grande distance alors que les conidies sont impliquées dans une dispersion à très courte distance (principalement à l'échelle de la plante) (Burt *et al.* 1997, Rutter *et al.* 1998, Burt 2003). Cependant, nos résultats apportent certains éléments nouveaux. Concernant la dispersion des conidies, nos résultats indiquent que la dispersion peut également survenir de façon rare à l'échelle d'une dizaine de mètres (jusqu'à 12.5 mètres sur le dispositif conidies de l'expérimentation directe). Néanmoins, les enseignements majeurs de cette thèse sont reliés à la mesure des capacités de dispersion des ascospores. En effet, jusqu'alors, il était supposé que la distance moyenne de dispersion des ascospores était de l'ordre d'une petite dizaine de mètres (voir pour exemple Abadie *et al.* in prep). L'ensemble des résultats obtenus dans cette thèse a permis de montrer que la distance moyenne de dispersion des ascospores était supérieure, plutôt de l'ordre de quelques centaines de mètres. Il était également supposé que la dispersion rare d'ascospores sur de longues distances était possible (Burt *et al.* 1998, Burt 2002, Amil 2007) mais cela n'avait pas pu être clairement démontré. Au cours de cette thèse, nous avons montré que la fonction de dispersion des ascospores est à queue lourde, c'est-à-dire que les événements de dispersion à longue distance sont plus fréquents que dans le cas d'une décroissance de type exponentielle. Nos résultats indiquent également que les ascospores peuvent disperser sur une distance d'au moins 1Km. En effet, cette distance étant la distance maximale à laquelle nous avons suivi la dispersion de façon directe, il est fort probable que d'autres ascospores aient pu disperser sur des distances plus importantes. La finalisation des analyses statistiques sur le jeu de données issu de l'expérimentation directe pourra probablement permettre de prédire les distances maximales attendues à partir de la distribution observée.

3- Implications

Une accélération des vitesses de colonisation

Les différences de capacités de dispersion observées entre les ascospores et les conidies apparaissent importantes à prendre en compte pour prédire l'évolution spatio-temporelle du pathogène dans des habitats non-colonisés ou en cours de colonisation. En effet, on peut imaginer que lorsque *M. fijiensis* arrive dans un nouvel environnement, il risque de commencer à se propager principalement grâce à la dispersion des conidies puisque la production d'ascospores nécessite des

densités de population suffisamment importantes pour permettre à deux lésions de signes sexuels compatibles de se rencontrer et de fusionner sur une même feuille de bananier (effet Allee¹). Il est donc possible que la vitesse de la vague d'expansion du pathogène soit dans un premier temps assez lente puisque la dispersion des conidies s'exerce sur de très courtes distances. Dans un second temps, on peut s'attendre à ce que la dispersion des ascospores générant une grande quantité de migrants longue distance entraîne une accélération des vitesses de colonisation (Kot *et al.* 1996). Les deux noyaux de dispersion mesurés pendant cette thèse pourront être intégrés dans des modèles épidémiologiques (basé sur leur convolution) afin préciser ce genre de prédictions. Dans un tel contexte, l'arrivée récente de *M. fijiensis* en Martinique (Octobre 2010), bien que très problématique pour la filière de production bananière, offre une situation très intéressante pour tester *in natura* ce genre de prédictions.

Un impact potentiel sur l'évolution des résistances aux fongicides

Le suivi temporel de la résistance aux Strobilurines et aux Benzimidazoles à l'interface entre une zone non-traitée et une zone actuellement traitée avec des fongicides de contact nous a permis de montrer que des niveaux significatifs de résistances associés à ces deux fongicides systémiques pouvaient persister pendant un grand nombre de générations après l'arrêt de la pression de sélection. Nous avons tout de même mis en évidence une diminution de la fréquence des allèles conférant la résistance à ces deux fongicides dans la zone industrielle (article 4). Puisque les fongicides systémiques en question ne sont plus utilisés, cette baisse des niveaux de résistance est théoriquement attendue et peut s'expliquer par l'effet de la contre-sélection et/ou des flux géniques en provenance des plantations non-traitées avoisinantes (i.e., phénomène de dilution). Nos résultats, obtenus à partir de modèles théoriques simplifiés visant à indépendamment estimer les valeurs de paramètres, ne nous ont pas permis de différencier l'importance relative de ces deux forces. Tout d'abord, les estimations réalisées en faisant abstraction de la migration semblent suggérer que s'il existe un coût en valeur sélective associé à la résistance des Strobilurines ou des Benzimidazoles, sa valeur est probablement faible. Ensuite, en considérant un modèle de migration unilatérale (de la zone non-traitée vers la zone traitée), nous avons montré que des taux de migration réalistes pouvaient également expliquer la baisse de fréquence observée en zone traitée. Les paramètres de dispersion estimés grâce à l'expérimentation directe confirment qu'une très forte asymétrie dans les flux géniques est attendue entre la zone non traitée et celle traitée aux fongicides de contacts. En effet, la zone non traitée présente une forte densité apparente, incluant la présence de nombreuses nécroses foliaires à l'origine de la production des ascospores. Nous avons montré que ces ascospores pouvaient être dispersées sur plusieurs centaines de mètres. Par opposition, la zone traitée aux fongicides de contacts présente une densité réduite et une quasi absence de nécroses foliaires. Seules des conidies présentant de faibles capacités de dispersion subsistent dans cette zone. Cependant, la distance maximale de dispersion mesurée lors de l'expérimentation directe (12.5m) indique que l'existence de flux géniques de la zone traitée vers la zone non-traitée est probablement faible mais ne peut pas être exclue puisque la distance séparant les deux

¹ L'effet Allee décrit un scénario dans lequel des populations à effectifs réduits sont affectées par une relation positive entre le taux d'accroissement et la densité.

compartiments sur le site d'étude est d'environ 10 mètres. La quantification des flux géniques dans cette direction apparaît donc importante puisque ces derniers peuvent réalimenter les zones protégées en gènes de résistance et de ce fait induire leur diffusion spatiale à large échelle par dispersion d'ascospores.

Bien que les résultats obtenus ne nous permettent pas de déterminer l'influence relative de la migration et de la sélection dans la diminution des niveaux de résistances observés sur le site d'étude, des données extérieures obtenues dans le cadre du projet global Européen dans lequel s'insère cette thèse peuvent s'avérer intéressantes à discuter. En effet, les niveaux de résistance aux Benzimidazoles ont été mesurés en 2004 et 2009 au sein de cinq plantations industrielles différentes de la zone de production bananière du Mounjo. La nature des prélèvements réalisés pour ces suivis n'est pas comparable à celle ayant permis l'étude présentée dans l'article 4 puisqu'un seul site a été échantillonné au sein des différentes plantations traitées aux fongicides de contacts. Cependant, bien que les niveaux de résistances aient diminué dans les cinq situations (Tableau 4), les estimations de coûts associés réalisés grâce au modèle théorique présenté dans l'article 4 (équation 4) reflètent des valeurs contrastées (pouvant aller jusqu'à un facteur ~6 entre le site 3 et 4). Une telle hétérogénéité dans ces valeurs de coûts mesurées laisse penser que la sélection n'est pas la seule force en jeu (car son intensité est censée être la même sur les différents sites) et que la migration des zones non-traitées vers les zones traitées est fort probablement impliquée dans la dilution des niveaux de résistances. La variation entre les valeurs mesurées sur les différents sites pourrait alors s'expliquer par des différences dans la proximité et la densité des zones non-traitées avoisinantes (à l'origine de patrons de flux géniques en direction des plantations industrielles d'intensité contrastées).

	Site d'étude (article 4)	Site annexe 1	Site annexe 2	Site annexe 3	Site annexe 4	Site annexe 5
p_0 (2004)	0.25	0.5	0.38	0.28	0.51	0.41
p (2009)	0.04	0.22	0.04	0.001	0.28	0.053
t	100	120	122	128	132	124
s	-0.0208	-0.0105	-0.0220	-0.0466	-0.0075	-0.0203

Tableau 4. Evolution temporelle des fréquences de résistance aux Benzimidazoles dans cinq autres plantations de la zone de production du Mounjo. Les valeurs de coût (s) ont été calculées grâce à l'équation suivante : $s = \frac{1}{t} \ln \frac{p(1-p_0)}{p_0(1-p)}$ (voir équation 4 de l'article 4 pour les détails sur l'équation)

Nous avons tout de même utilisé la théorie des clines pour décrire les patrons spatiaux de fréquences des allèles de résistance mais nous n'avons pas pu aller jusqu'à l'inférence de paramètres reliés au coût et à la migration par cette approche. La finalisation de ces analyses m'apparaît être une perspective d'étude importante à accomplir dans un futur proche. Pour cela, un nouvel échantillonnage a été réalisé en avril 2011 sur les mêmes sites que ceux considérés en janvier 2010. De plus, des données récoltées dans le but d'estimer la différence de densité des populations entre une zone traitée aux fongicides de contacts et une zone non-traitée sont en cours d'analyse. L'ensemble de ces données, combinées avec le développement d'un modèle théorique de

clines évoluant sous l'effet conjoint de la migration (bi latérale entre les deux zones) et d'un coût de la résistance pourront je l'espère permettre l'estimation des différents paramètres.

L'objectif du projet global dans lequel s'insère cette thèse est de déterminer s'il est possible de définir des stratégies efficaces (i.e., permettant un contrôle acceptable de la maladie) et durables basées sur la réutilisation des fongicides systémiques au sein des plantations industrielles de la zone de production bananière du Mounjo. Dans un tel contexte, une approche théorique de modélisation de l'évolution spatio-temporelle des résistances est nécessaire et les paramètres de dispersion mesurés pendant cette thèse devraient être utiles dans la paramétrisation de ces modèles. Il existe deux grands types de stratégies en termes de modélisation de la résistance aux fongicides et aux pesticides d'une façon plus générale (Van den Bosch & Gilligan 2008, Gilligan 2008). La première consiste à essayer de retarder l'apparition ou l'augmentation en fréquence des gènes de résistance alors que la seconde repose sur le maintien de la fréquence de ces gènes de résistance à une fréquence stable et acceptable. Dans l'optique du projet, la seconde stratégie apparaît plus adaptée puisque nous avons montré que des niveaux de résistances aux fongicides systémiques subsistaient dans les populations en l'absence de l'utilisation de ces fongicides. Cependant, l'existence de faibles fréquences d'allèles de résistance est un prérequis à l'application de cette stratégie de gestion des résistances. Dans ce contexte, il apparaît envisageable de « forcer » les flux géniques en direction des plantations industrielles dans le but d'accélérer ce phénomène de dilution des gènes de résistances. Pour cela, il serait envisageable d'installer au sein des plantations industrielles les mêmes structures artificielles mise en place dans l'expérimentation directe pour augmenter les flux génique à partir de tissus foliaires nécrosés prélevés en zones non traitées.

Le principe de la stratégie de maintien des niveaux de résistances est issu de modèles théoriques en génétique des populations visant à étudier la persistance d'un allèle dans un environnement spatialement hétérogène en termes de pression de sélection (Haldane 1948, Slatkin 1973, Nagylaki 1975, May 1975). Appliqué au cas de la résistance aux pesticides, le résultat principal de ces modèles indique que pour une population de taille donnée, il existe une fraction seuil de cette population au sein de laquelle on doit appliquer le traitement, fraction en dessous de laquelle la migration des individus de la zone non-traitée vers la zone traitée empêche l'allèle résistant d'envahir la population (Lenormand & Raymond 1998, Lenormand 2002). Cette fraction dépend du rapport des coefficients de sélection (sélection positive et coût) et de densités entre les zones traitées et non-traitées ainsi que de l'intensité de la migration entre les deux zones. Ainsi, l'existence d'un coût à la résistance est indispensable au confinement des niveaux de résistances sous une certaine valeur. En l'absence de coût, les allèles de résistances sont voués à la fixation lorsque les fongicides sont utilisés et seule une stratégie basée sur l'alternance temporelle de différents fongicides est envisageable.

Associés aux paramètres de migration mesurés pendant cette thèse, les paramètres de sélection actuellement étudiés dans une autre thèse pourront permettre de paramétrer un tel modèle. Pour la migration, il apparaît important de considérer indépendamment la dispersion des ascospores et des conidies puisque nous avons montré que ces deux types de spores présentent des capacités de dispersion bien différentes. Ainsi la migration pourrait être modélisée grâce aux noyaux de

dispersion des spores mesurés pendant cette thèse ce qui permettrait de considérer un modèle spatialement explicite. Grâce à une telle approche, il serait intéressant de commencer par calculer la taille de la zone critique (seuil) permettant de conserver un niveau acceptable de fréquences d'allèles résistants en zones industrielles (i.e., induisant une quantité de dégâts économiquement admissible par les producteurs). Cette taille permettrait de déterminer l'échelle spatiale à laquelle une stratégie de confinement des résistances doit se mettre en place. Le rapport entre cette taille et la taille globale d'une plantation industrielle sur la zone d'étude devrait donner une idée du caractère réaliste d'une telle stratégie de confinement. En conclusion, ce travail de modélisation apparaît intéressant mais requiert l'estimation des paramètres de sélection (actuellement en cours) pour pouvoir aboutir.

D'une façon plus théorique, les modèles de génétique des populations visant à calculer la taille d'une zone critique empêchant l'envahissement d'un ou plusieurs gène(s) de résistance ont toujours considéré un taux de migration ou bien une distance moyenne (σ^2). L'influence de différentes formes de noyaux sur les patrons d'évolution des résistances n'a à ma connaissance jamais été testée. Dans un contexte de modèles explicitement spatialisés, un tel développement pourrait s'avérer très intéressant à réaliser.

4 - Conclusions

L'ensemble des résultats obtenus pendant cette thèse indique que l'échelle géographique locale de l'agrosystème est la bonne pour mener une étude visant à caractériser les flux géniques issus de la dispersion naturelle de *M. fijiensis*. Dans un tel contexte, l'étude de la structuration génétique spatiale des populations s'avère être importante même si elle semble inadaptée à la caractérisation des flux géniques contemporains dans le cadre de populations non à l'équilibre génétique. En effet dans le cas présent, la détection d'une rupture spatiale dans les fréquences alléliques d'origine historique a été à l'origine du développement de l'approche basée sur les clines neutres et donc à celle de l'inférence d'un paramètre de migration à partir d'une méthode génétique. De plus, la stratégie basée sur la combinaison d'approches indirectes avec une méthode directe s'est également avérée fructueuse pour caractériser le processus de dispersion dans son ensemble. En effet, la cohérence entre les deux estimations nous procure une confiance supplémentaire envers les valeurs de paramètres inférés. Finalement, nos résultats ne permettent pas de réaliser des prédictions directes sur les patrons d'évolution des résistances aux fongicides systémiques mais ils seront probablement utiles pour paramétrer des modèles théoriques visant à déterminer si des stratégies de confinement des résistances aux fongicides systémiques apparaissent réalistes à mettre en place à l'échelle de la zone de production Camerounaise du Moungo.

Références bibliographiques

- Abadie C, Bonnot F, Lannou C, *et al.* (in prep) Measurement of the ascospores dispersal of the fungal pathogen of bananas *Mycosphaerella fijiensis* at a local scale.
- Abdo Z, Crandall KA, Joyce P (2004) Evaluating the performance of likelihood methods for detecting population structure and migration. *Molecular Ecology* **13**, 837-851.
- Agrios GN (2005) *Plant Pathology, Fourth Edition*.
- Amil AF, Heaney SP, Stanger C, Shaw MW (2007) Dynamics of QoI sensitivity in *Mycosphaerella fijiensis* in Costa Rica during 2000 to 2003. *Phytopathology* **97**, 1451-1457.
- Anderson PK, Cunningham AA, Patel NG, *et al.* (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**, 535-544.
- Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution* **24**, 21-30.
- Austerlitz F, Dick CW, Dutech C, *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* **13**, 937-954.
- Austerlitz F, Garnier-Gere PH (2003) Modelling the impact of colonisation on genetic diversity and differentiation of forest trees: interaction of life cycle, pollen flow and seed long-distance dispersal. *Heredity* **90**, 282-290.
- Austerlitz F, JungMuller B, Godelle B, Gouyon PH (1997) Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology* **51**, 148-164.
- Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics* **161**, 355-363.
- Awise JC, Arnold J, Ball RM, *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 459-522.
- Balkenhol N, Gugerli F, Cushman SA, *et al.* (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* **24**, 455-463.
- Banke S, McDonald BA (2005) Migration patterns among global populations of the pathogenic fungus *Mycosphaerella graminicola*. *Molecular Ecology* **14**, 1881-1896.
- Barres B, Carlier J, Seguin M, *et al.* (2011) Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *soumis*.
- Barton N (1979) Gene flow past a cline. *Heredity* **43**, 333-339.
- Barton N, Hewitt G (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.
- Barton NH (1983) Multilocus clines. *Evolution* **37**, 454-471. .
- Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process*. (ed. eds RGH). Oxford University Press, Oxford.
- Bazykin AD (1969) Hypothetical mechanism of speciation. *Evolution* **23**, 685-687.
- Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013-2029.

- Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology* **13**, 551-561.
- Bialozyt R, Ziegenhagen B, Petit RJ (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology* **19**, 12-20.
- Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology* **19**, 3515-3531.
- Birch PRJ, Whisson SC (2001) Phytophthora infestans enters the genomics era. *Molecular plant pathology* **2**, 257-263.
- Bohonak AJ (1999) Dispersal, gene flow, and population structure. *Quarterly Review of Biology* **74**, 21-45.
- Boileau MG, Hebert PDN, Schwartz SS (1992) Non-equilibrium gene frequency divergence : persistent founder effects in natural populations. *Journal of Evolutionary Biology* **5**, 25-39.
- Bossart JL, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution* **13**, 202-206.
- Brent KJ, Hollomon DW (1998) Fungicide resistance: The assessment of risk. *FRAC Monograph No. 2*.
- Breuillin F, Dutech C, Robin C (2006) Genetic diversity of the Chestnut blight fungus *Cryphonectria parasitica* in four French populations assessed by microsatellite markers. *Mycological Research* **110**, 288-296.
- Broquet T, Petit EJ (2009) Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology Evolution and Systematics* **40**, 193-216.
- Brown JKM (1997) *Plant pathogens and plant diseases* Australasian Plant Pathology Society Inc.
- Brown JKM, Hovmoller MS (2002) Epidemiology - Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537-541.
- Brunner PC, Schurch S, McDonald BA (2007) The origin and colonization history of the barley scald pathogen *Rhynchosporium secalis*. *Journal of Evolutionary Biology* **20**, 1311-1321.
- Bullock JM, Shea K, Skarpaas O (2006) Measuring plant dispersal: an introduction to field methods and experimental design. *Plant Ecology* **186**, 217-234.
- Burt PJA (2003) Airborne dispersal of *Mycosphaerella fijiensis*. In: *Mycosphaerella Leaf Spot Diseases of Bananas: Present Status and Outlook. Proceedings of the Workshop on Mycosphaerella Leaf Spot Diseases, San José, Costa Rica*.
- Burt PJA, Rutter J, Gonzales H (1997) Short-distance wind dispersal of the fungal pathogens causing Sigatoka diseases in banana and plantains. *Plant Pathology* **46**, 451-458.
- Burt PJA, Rutter J, Ramirez F (1998) Airborne spore loads and mesoscale dispersal of the fungal pathogens causing Sigatoka diseases in banana and plantain. *Aerobiologia* **14**, 209-214.
- Calmet C (2002) *Inférences sur l'histoire des populations à partir de leur diversité génétique: étude de séquences démographiques de type fondation-explosion*, Université Paris VI.
- Carlier J, Lebrun MH, Zapater MF, Dubois C, Mourichon X (1996) Genetic structure of the global population of banana black leaf streak fungus, *Mycosphaerella fijiensis*. *Molecular Ecology* **5**, 499-510.
- Churchill ACL (2010) *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Molecular plant pathology* **12**, 307-328.

- Clobert (2001) *Dispersal*, Clobert J, Danchin E, Dhondt AA, Nichols JD editions edn. New York: Oxford Univ. Press.
- Cockerham CC, Weir BS (1993) Estimation of gene flow from F-statistics. *Evolution* **47**, 855-863.
- Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.
- Cox KD, Scherm H (2001) Gradients of primary and secondary infection by *Monilinia vaccinii-corymbosi* from point sources of ascospores and conidia. *Plant Disease* **85**, 955-959.
- De Lapeyre de Bellaire L., Fouré E., Abadie C., Carlier J (2010) Black leaf streak disease is challenging the banana industry. *Fruits* **65**, 327-342.
- Dieckmann U, O'Hara B, Weisser W (1999) The evolutionary ecology of dispersal. *Trends in Ecology & Evolution* **14**, 88-90.
- Dirienzo A, Peterson AC, Garza JC, *et al.* (1994) Mutational processes of simple-sequence repeat loci in human-populations. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3166-3170.
- Dobzhansky T, Wright S (1943) Genetics of natural populations. X. Dispersal rates in *Drosophila pseudoobscura*. *Genetics* **28**, 304-340.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 975-979.
- Endler JA (1977) Geographic variation, speciation, and clines. *Monographs in Population Biology* **10**, 1-246.
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* **19**, 4113-4130.
- Et-touil K, Bernier L, Beaulieu J (1999) Genetic structure of *Cronartium ribicola* populations in eastern Canada. *Phytopathology* **89**, 915-919.
- Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology Evolution and Systematics* **40**, 481-501.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
- Fayard J, Klein EK, Lefevre F (2009) Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology* **22**, 2171-2182.
- Fisher RA (1930) *The genetical theory of natural selection*, Clarendon Press, Oxford. edn. Clarendon Press, Oxford. .
- Fontem DA, Berger RD, Weingartner DP, Bartz JA (1991) Progress and spread of dark leaf-spot in cabbage. *Plant Disease* **75**, 269-274.
- Fouré E (1982) Les cercosporioses du bananier et leurs traitements. Comportement des variétés. Etude de la sensibilité variétale des bananiers et plantains à *Mycosphaerella fijiensis* Morelet au Gabon (maladie des raies noires). *Fruits* **40**, 393-399.

- Fouré E (1988) Stratégies de lutte contre la cercosporiose noire des bananiers et des plantains provoquée par *Mycosphaerella fijiensis* Morelet. L'avertissement biologique au Cameroun. Evaluation des possibilités d'amélioration. *Fruits* **43**, 269-274.
- Fouré E, Lescot T (1988) Variabilité génétique des *Mycosphaerella* inféodés au genre *Musa*. Mise en évidence de la présence au Cameroun sur bananiers et plantains d'une cercosporiose (*Mycosphaerella musicola*) au comportement pathogène atypique. *Fruits* **43**, 407-415.
- Francois O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* **174**, 805-816.
- Francois O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* **10**, 773-784.
- Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **176**, 1635-1651.
- Gauhl F (1994) Epidemiology and ecology of black sigatoka (PhD Thesis), p. 120. INIBAP, Montpellier.
- Gauhl F, Pasberg-Gauhl C, Jones DR (2000) Black leaf streak: disease cycle and epidemiology. In: *Diseases of Banana*, (ed. Abacá and Enset (Jones DR, ed.)), pp. 56-62. New York: CABI Publishing.
- Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis*.
- Gilligan CA (2008) Sustainable agriculture and plant diseases: an epidemiological perspective. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 741-759.
- Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring Population Decline and Expansion from Microsatellite Data: a Simulation-Based Evaluation of the MSVAR Method. *Genetics Advance Online Publication*.
- Gladieux P, Zhang XG, Afoufa-Bastien D, *et al.* (2008) On the Origin and Spread of the Scab Disease of Apple: Out of Central Asia. *Plos One* **3**.
- Gobbin D, Rumbou A, Linde C, Gessler C (2006) Population genetic structure of *Plasmopara viticola* after 125 years of colonization in European vineyards. *Molecular plant pathology* **7**, 519-531.
- Guerin F, Gladieux P, Le Cam B (2007) Origin and colonization history of newly virulent strains of the phytopathogenic fungus *Venturia inaequalis*. *Fungal Genetics and Biology* **44**, 284-292.
- Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation-the correlated allele frequencies model revisited. *Bioinformatics* **24**, 2222-2228.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.
- Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology* **18**, 4734-4756.
- Guillot G, Santos F (2009) A computer program to simulate multilocus genotype data with spatially autocorrelated allele frequencies. *Molecular Ecology Resources* **9**, 1112-1120.
- Haldane JBS (1948) The theory of a cline. *Journal of Genetics* **48**, 277-284.

- Halkett F, Coste D, Rivas Platero GG, *et al.* (2010) Isolation by distance and genetic discontinuities in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Molecular Ecology*.
- Hallatschek O, Hersen P, Ramanathan S, Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19926-19930.
- Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145-154.
- Hartl D, Clark A (2007) *Principles of Population Genetics, Fourth Edition*.
- Hastings WK (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 97-109.
- Hayden HL, Carlier J, Aitken EAB (2003a) Genetic structure of *Mycosphaerella fijiensis* populations from Australia, Papua New Guinea and the Pacific Islands. *Plant Pathology* **52**, 703-712.
- Hayden HL, Carlier J, Aitken EAB (2003b) Population differentiation in the banana leaf spot pathogen *Mycosphaerella musicola*, examined at a global scale. *Plant Pathology* **52**, 713-719.
- Hayden HL, Cozijnsen AJ, Howlett BJ (2007) Microsatellite and minisatellite analysis of *Leptosphaeria maculans* in Australia reveals regional genetic differentiation. *Phytopathology* **97**, 879-887.
- Henderson J, Pattemore JA, Porchun SC, *et al.* (2006) Black Sigatoka disease: new technologies to strengthen eradication strategies in Australia. *Australasian Plant Pathology* **35**.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.
- Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetic research* **38**, 226-231.
- Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* **58**, 199-207.
- Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**, 282-291.
- Inman AJ, Fitt BDL, Todd AD, Evans RL (1999) Ascospores as primary inoculum for epidemics of white leaf spot (*Mycosphaerella capsellae*) in winter oilseed rape in the UK. *Plant Pathology* **48**, 308-319.
- Jeger MJ (1989) The spatial component of plant diseases epidemics. *ed. Prentice-Hall Inc., Englewood, Cliffs, NJ.*, Pages 1-13.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *Bmc Genetics* **11**.
- Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**, 92-103.
- Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology* **12**, 2511-2523.
- Jones DR (2000) Diseases of banana, abaca and enset. CABI Publishing, Oxon.

- Kass RE, Raftery AE (1995) Bayes Factor. *Journal of the American Statistical Association* **90**, 773-795.
- Klein EK, Lavigne C, Picault H, Renard M, Gouyon PH (2006) Pollen dispersal of oilseed rape: estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology* **43**, 141-151.
- Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* **23**, 482-490.
- Knowles LL (2009) Statistical Phylogeography. *Annual Review of Ecology Evolution and Systematics* **40**, 593-612.
- Koenig WD, VanVuren D, Hooge PN (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology & Evolution* **11**, 514-517.
- Kot M, Lewis MA, vandenDriessche P (1996) Dispersal data and the spread of invading organisms. *Ecology* **77**, 2027-2042.
- Labbé P (2005) *Evolution de l'adaptation locale: Le cas de la résistance aux insecticides chez le moustique Culex pipiens*, PhD thesis Université de Montpellier 2 - France.
- Landguth EL, Cushman SA, Schwartz MK, et al. (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* **19**, 4179-4191.
- Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a "Continuous" population under isolation by distance. *Molecular Biology and Evolution* **20**, 491-502.
- Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* **166**, 1081-1092.
- Lebreton JD, Hines JE, Pradel R, Nichols JD, Spendelov JA (2003) Estimation by capture-recapture of recruitment and dispersal over several sites. *Oikos* **101**, 253-264.
- Lenormand T (1998) *Dynamique de l'adaptation locale: Résistance aux insecticides chez Culex pipiens*, PhD thesis Ecole Nationale Supérieure Agronomique de Montpellier
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**, 183-189.
- Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* **400**, 861-864.
- Lenormand T, Guillemaud T, Bourguet D, Raymond M (1998) Evaluating gene flow using selected markers: A case study. *Genetics* **149**, 1383-1392.
- Lenormand T, Raymond M (1998) Resistance management: the stable zone strategy. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 1985-1990.
- Lenormand T, Raymond M (2000) Analysis of clines with variable selection and variable migration. *American Naturalist* **155**, 70-82.
- Lepoivre (2003) *Phytopathologie*, De boeck edn.
- Lescot T (2006) La banane en chiffres : le fruit préféré de la planète. *Fruitrop* **140**, 5-9.
- Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic-structure of a tropical understory shrub, *psychotria officinalis* (rubiaceae). *American Journal of Botany* **82**, 1420-1425.
- Luikart G, Cornuet JM (1999) Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics* **151**, 1211-1216.

- Mallet J, Barton N, Lamas G, *et al.* (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in heliconius hybrid zones. *Genetics* **124**, 921-936.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.
- Maruyama T, Fuerst PA (1985) Population bottlenecks and nonequilibrium models in population-genetics .2. number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**, 675-689.
- May RM, Endler JA, McMurtrie RE (1975) Gene frequency clines in the presence of selection opposed by gene flow. *The American Naturalist* **109**, 659-676.
- Mayr (1963) *Animal species and evolution*, Cambridge, MA: Harvard University Press, edn. Cambridge, MA: Harvard University Press, .
- Mayr (1970) *Populations, Species, and Evolution*. , The Belknap Press of Harvard University Press, Cambridge, MA edn. The Belknap Press of Harvard University Press, Cambridge, MA.
- McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349-+.
- McRae BH (2006) Isolation by resistance. *Evolution* **60**, 1551-1561.
- McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19885-19890.
- Medlock J, Kot M (2003) Spreading disease: integro-differential equations old and new. *Mathematical Biosciences* **184**, 201-222.
- Milgroom MG, Lipari SE, Wang KR (1992) Comparison of genetic diversity in the chestnut blight fungus, *Cryphonectria* (*Endothia*) *parasitica*, from China and the U.S. *Mycological Research* **96**, 1114-1120.
- Milgroom MG, Wang KR, Zhou Y, Lipari SE, Kaneko S (1996) Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* **88**, 179-190.
- Montarry J, Andrivon D, Glais I, *et al.* (2010) Microsatellite markers reveal two admixed genetic groups and an ongoing displacement within the French population of the invasive plant pathogen *Phytophthora infestans*. *Molecular Ecology* **19**, 1965-1977.
- Mourichon X, Carlier J, Fouré E (1997) Les cercosporioses: Maladie des raies noires (Cercosporiose noire); Maladie de Sigatoka (Cercosporiose jaune). In: *fiche technique n°8* (ed. Musa Md), p. 4p. INIBAP, Montpellier.
- Mundt CC (1989) Use of the modified Gregory model to describe primary disease gradients of wheat leaf rust produced from area sources of inoculum. *Phytopathology* **79**.
- Nagylaki T (1975) Conditions for the existence of clines. *Genetics*, 595-615.
- Nathan R, Schurr FM, Spiegel O, *et al.* (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution* **23**, 638-647.
- Nei M, Tajima F (1981) Genetic drift and estimation of effective population-size. *Genetics* **98**, 625-640.
- Nichols RA, Hewitt GM (1994) The genetic consequences of long-distance dispersal during colonization. *Heredity* **72**, 312-317.
- Oddou-Muratorio S, Klein EK (2008) Comparing direct vs. indirect estimates of gene flow within a population of a scattered tree species. *Molecular Ecology* **17**, 2743-2754.

- Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.
- Palm M (2001) Systematics and the Impact of Invasive Fungi on Agriculture in the United States. *Bioscience* **51**, 141-147.
- Parisod C, Bonvin G (2008) Fine-scale genetic structure and marginal processes in an expanding population of *Biscutella laevigata* L. (Brassicaceae). *Heredity* **101**, 536-542.
- Parnell M, Burt PJA, Wilson K (1998) The influence of exposure to ultraviolet radiation in simulated sunlight on ascospores causing Black Sigatoka disease of banana and plantain. *International Journal of Biometeorology* **42**, 22-27.
- Plummer M, Best N, Cowles K, Vines K (2006) Coda: output analysis and diagnostics for MCMC. *R News* **6**, 7-11.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Ramsey MD, Daniells JW, Anderson DJ (1990) Effects of Sigatoka leaf spot (*Mycosphaerella musicola* Leach) on fruit yields, field ripening and greenlife of bananas in North Queensland. *Sci. Hortic* **41**, 305-313.
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9197-9201.
- REX Consortium (2010) The skill and style to model the evolution of resistance to pesticides and drugs. *Evolutionary Applications* **3**, 375-390.
- Rhodes PL (1964) A new banana disease in Fiji. . *Commonwealth Phytopathology News* **10**, 38-41.
- Rivas GG, Zapater MF, Abadie C, Carlier J (2004) Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis*. *Molecular Ecology* **13**, 471-482.
- Robert S, Ravigne V, Zapater MF, C. A, Carlier J (2011) Contrasting introduction patterns among continents in the worldwide invasion of the banana fungal pathogen *Mycosphaerella fijiensis*. *soumis*.
- Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics* **173**, 1033-1045.
- Ronce O (2007) How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology Evolution and Systematics* **38**, 231-253.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.
- Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.
- Rousset F (2001) Genetic approaches to the estimation of dispersal rates. . In: *Dispersal*, pp. 18-28.
- Rousset F (2007) Inferences from spatial population genetics. . In: *Handbook of Statistical Genetics* (ed. ed. DJ Balding MB, C Cannings), pp. 945-979.
- Rousset F (2008) Dispersal estimation: Demystifying Moran's I. *Heredity* **100**, 231-232.
- Rutter J, Burt PJA, Ramirez. F (1998) Movement of *Mycosphaerella fijiensis* spores and Sigatoka disease development on plantain close to an inoculum source. *Aerobiologica* **14**, 201-208.

- Saccheri IJ, Rousset F, Watts PC, Brakefield PM, Cook LM (2008) Selection and gene flow on a diminishing cline of melanic peppered moths. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16212-16217.
- Scherin H (1996) On the velocity of epidemic waves in model plant disease epidemics. *Ecological Modelling* **87**, 217-222.
- Sharrock S, Frison E (1999) *Musa production around the world - Trends, varieties and regional importance*, INIBAP Annual Report edn. INIBAP Annual Report.
- Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733-756.
- Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics*, 396-430.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.
- Slatkin M, Barton N (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43**, 1349-1368.
- Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology & Evolution* **14**, 219-224.
- Storfer A, Murphy MA, Evans JS, *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128-142.
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology* **19**, 3496-3514.
- Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* **56**, 154-166.
- Stover RH (1972) *Banana, plantain, and abaca diseases*, Commonwealth Mycological Institute edn. Commonwealth Mycological Institute.
- Stover RH (1983) Effet du cercospora noir sur les plantains en Amérique centrale. *Fruits* **38**, 326-329.
- Sumner J, Rousset F, Estoup A, Moritz C (2001) 'Neighbourhood' size, dispersal and density estimates in the prickly forest skink (*Gnypetoscincus queenslandiae*) using individual genetic and demographic methods. *Molecular Ecology* **10**, 1917-1927.
- Templeton AR, H. Hemmer, G. Mace, U.S. Seal, W.M. Shields, and D.S. Woodruff. (1986) Local adaptation, coadaptation, and population boundaries. *Zoo Biol* **5**, 115-125.
- Van den Bosch F, Gilligan CA (2008) Models of fungicide resistance dynamics. *Annual Review of Phytopathology* **46**, 123-147.
- Vandewoestijne S, Baguette M (2004) Demographic versus genetic dispersal measures. *Population Ecology* **46**, 281-285.
- Vitalis R, Couvet D (2001) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics* **157**, 911-925.
- Wang JL (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B-Biological Sciences* **360**, 1395-1409.
- Waples R, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

- Waples RS (1989) A generalized-approach for estimating effective population-size from temporal changes in allele frequency. *Genetics* **121**, 379-391.
- Watts PC, Rousset F, Saccheri IJ, *et al.* (2007) Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion mercuriale* : Odonata : Zygoptera) populations: analysis of 'neighbourhood size' using a more precise estimator. *Molecular Ecology* **16**, 737-751.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
- Williamson EG, Slatkin M (1999) Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics* **152**, 755-761.
- Wilmer J, Elkin C, Wilcox C, *et al.* (2008) The influence of multiple dispersal mechanisms and landscape structure on population clustering and connectivity in fragmented artesian spring snail populations. *Molecular Ecology* **17**, 3733-3751.
- Wilson GF, Buddenhagen I (1986) The black sigatoka threat to plantain and banana in West Africa. *IITA Research Highlights* **7**.
- Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* **150**, 499-510.
- Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 97-159.
- Wright S (1943) Isolation by Distance. *Genetics*, 114-138.
- Yafetto L, Carroll L, Cui YL, *et al.* (2008) The Fastest Flights in Nature: High-Speed Spore Discharge Mechanisms among Fungi. *Plos One* **3**.

Document annexe n°3 : Mise au point d'une méthode d'extraction d'ADN directement à partir de lésions et développement de nouveaux marqueurs microsatellites chez *Mycosphaerella fijiensis*.

Je présente ici la note technique publiée dans *American Journal of Botany* décrivant certaines des mises au point réalisées en début de thèse conjointement avec Stéphanie Robert (Thèse CIRAD/Région). Deux améliorations ont été apportées dans le but d'optimiser la technique du génotypage moléculaire chez *M. fijiensis*. La première concerne l'extraction d'ADN directement à partir de lésions présentes sur les feuilles échantillonnées sur le terrain. En effet, la technique précédemment employée (utilisée dans l'article n°1) était basée sur l'isolement au laboratoire des spores du champignon à partir de feuilles de bananiers infectées et leur culture sur milieu nutritif gélosé. Outre son caractère laborieux, cette étape était caractérisée par un taux de perte et/ou de contamination non négligeable. La méthode mise au point nous a donc permis d'envisager des échantillonnages de tailles plus importantes. La seconde amélioration concerne le développement d'une vingtaine de nouveaux marqueurs microsatellites (à motifs tétra-nucléotidiques) pouvant être amplifiés par multiplexage et permettant de combiner différentes études réalisées à des échelles spatiales diverses.

OPTIMIZED GENOTYPING WITH MICROSATELLITE MARKERS IN THE FUNGAL BANANA PATHOGEN *MYCOSPHAERELLA FIJIENSIS* (MYCOSPHAERELLACEAE)¹

STÉPHANIE ROBERT², ADRIEN RIEUX², XAVIER ARGOUT³, JEAN CARLIER^{2,4},
AND MARIE-FRANÇOISE ZAPATER²

²CIRAD, Unité Mixte de Recherche–Biologie et Génétique des Interactions Plante-Parasite, TA A-54/K, Campus International de Baillarguet, F-34398 Montpellier Cedex 5, France; ³CIRAD, Unité Mixte de Recherche–Développement et Amélioration des Plantes, TA A-96/03, Avenue Agropolis, F-34398 Montpellier Cedex 5, France

- *Premise of the study:* Large-scale population genetics studies are required to investigate the dispersal processes underlying the emergence of *Mycosphaerella fijiensis*, a fungal pathogen of banana. To this end, we have developed an optimized genotyping procedure combining novel microsatellite markers and a modified DNA extraction protocol.
- *Methods and Results:* Primers for tetranucleotide loci were designed directly from the recently published genome sequence of *M. fijiensis*. A total of 19 new polymorphic and easy-to-score markers were developed. Their use was combined with an adapted protocol for total DNA extraction starting from young lesions collected from banana leaves, thus avoiding a pathogen isolation step.
- *Conclusions:* The combination of the two technical developments presented here will permit the expansion of genotyping capacity in *M. fijiensis*, allowing large-scale analysis of samples from various geographic locations.

Key words: DNA extraction; genotyping; *Mycosphaerella fijiensis*; tetranucleotide microsatellites.

Population genetics studies on fungi are crucial to our understanding of the dispersal process of such organisms causing emerging diseases in plants. Such studies can be extended by integrating tools from landscape genetics (Archie et al., 2009). These tools require a sufficient number of neutral genetic markers and increasingly large genotyping capacities. For plant-pathogenic fungi, an important advantage is that DNA can be extracted directly from plant lesions, avoiding time-consuming isolation of the pathogen. However, this technique has seldom been used and needs to be combined with specific, easy-to-score markers to compensate for the possibly suboptimal amplification conditions.

The Ascomycete and haploid fungus *Mycosphaerella fijiensis*—the causal agent of the black leaf streak disease of bananas—can be considered as a model in the field of emerging plant diseases (Halkett et al., 2010). It has spread recently to almost all banana-producing regions worldwide and is considered among the most serious biological threats to food security. Population structure studies with microsatellite markers have been published at local scales (Halkett et al., 2010). However, most of

the microsatellite loci previously published in *M. fijiensis* (Neu et al., 1999; Zapater et al., 2008; Yang and Zhong, 2008) have been adapted for specific genetic studies at local scales, limiting comparisons among localities at continental or global scales. Indeed, most such loci correspond to di- and trinucleotides or imperfect motifs, thus increasing errors in allele identification when the number of alleles is high, as would be expected at large geographical scales. Moreover, their use is based on DNA extraction from strains isolated and cultivated on artificial medium. Consequently, to address both these issues, we have developed new microsatellite markers corresponding to tetranucleotide motifs identified in the recently published genome sequence of *M. fijiensis*. The use of these easier-to-score markers was combined with a low-cost DNA extraction protocol starting from young lesions collected on banana leaves.

METHODS AND RESULTS

The nuclear genome sequence of *M. fijiensis* was released recently on the JGI (Joint Genome Institute, <http://genome.jgi-psf.org/Mycfi2/Mycfi2.home.html>); the sequence was analyzed for microsatellite content using the program MISA (MICROSatellite identification tool, <http://pgsc.ipk-gatersleben.de/misa>). We selected 35 tetranucleotide loci scattered throughout the genome. A total of 35 primer pairs was designed with PRIMER3 software (<http://frodo.wi.mit.edu/primer3/>, Rozen and Skaletsky, 2000), with an optimum size of 20 nucleotides, an optimum melting temperature of 60.0°C and a minimum GC content of 50%. These primer pairs were first tested on DNA from eight *M. fijiensis* isolates from Asia, Africa, and America following the procedure described in Zapater et al. (2008). About 20 loci giving rise to a single amplification product were selected on 2% agarose gels.

The polymorphism of these sequences was further tested on a global DNA sample of 110 cloned isolates from 19 countries around the world (including Taiwan, China, Vietnam, Philippines, Indonesia, Malaysia, Papua New Guinea,

¹ Manuscript received 6 August 2010; revision accepted 16 September 2010.

The authors thank Arnaud Estoup for helpful discussions and Helen Rothnie for her attentive reading and careful English-language checking. Data used in this work were produced through the molecular genetic analysis technical facilities of the IFR119 “Montpellier Environnement Biodiversité.” This work was supported by the région Languedoc Roussillon, Bayer CropScience, the Agence Nationale de la Recherche et de la Technologie, and the Agence Nationale de la Recherche, contract number ANR 07-BDIV-003 (Emerfundis project).

⁴ Author for correspondence: jean.carlier@cirad.fr

Costa Rica, Honduras, Panama, Colombia, Dominican Republic, Jamaica, Cuba, Cameroon, Ivory Coast, Nigeria, Uganda, and Comoros), and two populations of 40 isolates from Cameroon and Costa Rica. All the strains were isolated and conserved in the collection of CIRAD. Total DNA was extracted from cultivated mycelium in vitro as described in Halkett et al. (2010). PCR was performed according to the following multiplex protocol (adapted from Zapater et al., 2008): 15 min at 95°C, 45 cycles of 30 s at 94°C, 90 s at 57°C and 90 s at 72°C, and a 10-min final step at 65°C. The 10-μL reaction mixture included 3 μL DNA (5 ng/μL), 5 μL (2×) QIAGEN Multiplex PCR Master Mix (Qiagen, Valencia, California, USA), 1 μL (5×) Q-solution and 0.5 μL of each primer. Forward or reverse primers were labeled with one of the following fluorescent dyes: NED, HEX, FAM, or PET (Applied Biosystems, Foster City, California, USA). Primers were tested in panels of four markers labeled with those different dyes. Diluted amplified PCR products were mixed with GeneScan-500LIZ size marker (Applied Biosystems) and separated on a 16-capillary Sequencer (ABI Prism 3130XL, Applied Biosystems). They were sized and analyzed using GeneMapper Software (Applied Biosystems); 19 loci that were technically easy to analyze were retained (Table 1).

Tetranucleotide motif markers appear particularly easy to score because alleles were, as expected, separated by 4 bp and thus were easy to identify without ambiguity across samples.

The mean number of null alleles across loci ranged from 0.9% (1/110) to 6.4% (7/110) in the global sample and from 0 to 7.5% (3/40) in the two populations. Genetic diversity was estimated by the number of alleles observed (N_a)

and the expected heterozygosity (H_E), calculated using FSTAT_{2.9.3} (Goudet, 1995; Table 2). At the global scale for each locus, N_a ranged between 3 and 14 (mean 7.58), and H_E from 0.31 to 0.80 (mean 0.63). Analyses were also performed within the three distinct continents: Africa and America where *M. fijiensis* was recently introduced, and Southeast Asia corresponding to its center of origin. As expected, all markers were polymorphic in Southeast Asia, with a mean N_a and H_E across loci of 6.9 and 0.74, respectively. They displayed lower levels of polymorphism in Africa and America with a mean N_a of 2.11 and 1.95, and a mean H_E of 0.30 and 0.32, respectively. Within populations from Cameroon and Costa Rica, the mean N_a and H_E varied between 1.5 to 2 and 0.18 to 0.27, respectively. No significant linkage disequilibrium among any pair of loci was detected using Fisher's exact tests in the GENEPOP program (Raymond and Rousset, 1995), and we used the false discovery rate procedure in the QVALUE package (Storey, 2002) to control for multiple testing. All markers can easily be included in multiplex panels in the coming studies. Since some markers appeared monomorphic in some areas, different sets of loci will have to be chosen for microsatellite multiplexing, depending on the geographic area studied.

Simultaneously, we adapted the DNA extraction protocol starting from in vitro grown mycelium described in Halkett et al. (2010) for young lesions directly cut out from banana leaves. Individual lesions of stage 2 (see description in Jones, 2000) were cut and dried for 1 d at 55°C. In contrast with the protocol described in Halkett et al. (2010), the extraction buffer was added in a two-step procedure allowing a better grinding of the lesions.

TABLE 1. Characteristics of the 19 tetranucleotide microsatellite loci developed for *Mycosphaerella fijiensis*.

Locus name	Primer sequences	Repeat motif	Genome position ^a	Expected size (bp)	Allele size range ^b (bp)	Annealing temperature T_a (°C)
MfSSR401	F: GACGTGGCAACATCAACATC R: CTCAGCCTGTGGTGTGTTGAG	(CTCA) ₇	1/8304341–8304529	188	173–227	55
MfSSR402	F: CTTCTCGTCTCCACACTTCC R: CACACGAACCCCTTGTCATCA	(TTGC) ₇	5/1559726–1559952	226	215–231	55
MfSSR403	F: TAGTGGTTGGCGGATATGGT R: AGGCTCAGCTCATCGGTAGA	(CAAG) ₁₄	10/2377024–2377220	196	162–215	55
MfSSR404	F: GCTGCTTTGGTTTGTAAAG R: GTCGAACCTTGCCCTGAGAC	(CAGG) ₁₂	4/2435930–2436142	212	194–231	55
MfSSR405	F: TGTTCACGGAGCAAGACAAG R: GATGGTATGGCGTGGTCTCT	(CATT) ₆	12/824577–824693	116	101–120	55
MfSSR406	F: AGCCATTGTGTAGCGTCTTG R: GGCAATACAACCCGAGGATA	(CTTC) ₆	2/922431–922602	171	171–183	55
MfSSR407	F: CTCATATGGCTCAGGGTGAAT R: TGCTGGAAAGGAGACTTTGG	(CCAT) ₇	5/3498739–3498949	210	206–244	55
MfSSR410	F: GTAACGAACCTGTCTGCATA R: CCCCAATGTCGTAAACCATC	(CCAT) ₁₁	10/656074–656203	129	103–172	55
MfSSR412	F: CGGGGTCTCGGTATCAAGTA R: CGTGTTACGCTTGATTGTGG	(TTGC) ₇	15/603217–603389	172	145–187	55
MfSSR413	F: TGGGCCCTTACTTGTGAAC R: AACACGGGGGTACAAGACAG	(TCAC) ₇	7/2701378–2701597	219	201–245	55
MfSSR417	F: AGGGTGTACACATCGGCTTC R: CTTCTTGTGCAACCGCTAC	(TTCT) ₆	5/5244786–5244989	203	203–254	55
MfSSR418	F: GGACATGTCTGATTTGATGC R: TCGACCTGTACATCTCCAA	(GCCT) ₈	2/8232842–8233017	175	150–179	55
MfSSR420	F: CGAGCGAAGCTGAATAGTCC R: GTTCTGTGCTTGAGCGAGTG	(CTGA) ₆	7/3524840–3525030	190	182–202	55
MfSSR422	F: GAGCTGCAACCAAGAGAAGC R: GCTTCTAGCCCTGCAAACAC	(TCCA) ₆	6/1052203–1052412	209	196–213	55
MfSSR424	F: CGATACCCACCACATTGAT R: CTTCTGCTGAGAGCGGAATC	(TACC) ₅	2/4088730–4088903	173	175–194	55
MfSSR425	F: AACCACACATCTGTCTCC R: TTAGAGGTTGGATGCCGAGT	(TGAT) ₅	2/1374137–1374364	227	222–235	55
MfSSR428	F: AGGATTACCCGAGAACAG R: GGGTCTAGGGAATGCCTACT	(GAAG) ₅	4/5852974–5853108	134	127–131	55
MfSSR430	F: TAGCGAGGTGCATGGTACTG R: GGGGTTGAGGGATATTCGAC	(CTTT) ₅	7/4088786–4088968	182	170–186	55
MfSSR434	F: TAACGTGCGAGGAGGACTCT R: AGATCAGTGCCCCCTTGAGAA	(TACC) ₅	4/1802056–1802280	224	209–253	55

^a Scaffold/sequence position from the forward to the reverse primers according to the *M. fijiensis* genome sequence v2.0 (<http://genome.jgi-psf.org/Mycfi2/Mycfi2.home.html>). Amplified DNA sequences can be retrieved by copying their positions to the browser of the JGI site.

^b Based on a global DNA sample of 110 isolates.

F = forward primer, R = reverse primer

TABLE 2. Genetic diversity analyses at continental and intrapopulation scales in *Mycosphaerella fijiensis*, using the 19 tetranucleotide microsatellite markers.

Locus	Global scale (N = 110)						Intrapopulation			
	Southeast Asia (N = 32)		Africa (N = 45)		Latin America (N = 33)		Cameroon (N = 40)		Costa Rica (N = 40)	
	H _E	N _a	H _E	N _a	H _E	N _a	H _E	N _a	H _E	N _a
MfSSR401	0.83	8	0.24	2	0.22	2	0.49	2	0.14	2
MfSSR402	0.74	4	0	1	0.52	2	0	1	0.55	3
MfSSR403	0.85	7	0.22	5	0.48	2	0	1	0.51	2
MfSSR404	0.88	14	0.51	2	0.23	2	0.37	2	0.23	3
MfSSR405	0.85	9	0.51	2	0.52	3	0.48	2	0.63	3
MfSSR406	0.69	4	0.17	2	0.48	2	0	1	0.43	2
MfSSR407	0.84	8	0.51	2	0.47	2	0.51	2	0.50	2
MfSSR410	0.89	12	0.09	2	0.46	2	0	1	0.51	2
MfSSR412	0.86	8	0.51	2	0.31	2	0.19	2	0.41	2
MfSSR413	0.85	9	0.48	2	0.29	2	0.34	2	0.11	2
MfSSR417	0.77	6	0.51	2	0.62	4	0.51	2	0.52	4
MfSSR418	0.68	3	0.33	4	0.33	2	0.05	2	0	1
MfSSR420	0.79	6	0	1	0	1	0	1	0	1
MfSSR422	0.76	6	0	1	0	1	0	1	0	1
MfSSR424	0.83	8	0.51	2	0.52	2	0.51	2	0.10	2
MfSSR425	0.39	4	0.14	2	0.19	2	0	1	0.06	2
MfSSR428	0.41	3	0.31	2	0.49	2	0	1	0.47	2
MfSSR430	0.65	6	0.13	2	0	1	0	1	0	1
MfSSR434	0.59	5	0.50	2	0.00	1	0.05	2	0.05	2
Means	0.74	6.90	0.30	2.11	0.32	1.95	0.18	1.53	0.27	2.05

N = size of the sample analyzed, H_E = expected gene diversity, N_a = number of alleles

Dried lesions were first ground in 200 µl of 65°C extraction buffer (1.05 M NaCl, 75 mM Tris-HCl 1M pH 8.0, 15mM EDTA 0.5M pH 8.0) in 2 ml micro-centrifuge tubes, then 100 µl of 65°C cetyltrimethylammonium bromide (CTAB) buffer (3% CTAB, 1.5% sodium sulfite) was added and the tubes were placed at 65°C for 60 min. The other steps of the extraction protocol described in Halkett et al. (2010) remained unchanged.

A selection of 11 new microsatellite markers polymorphic at the scale of Cameroon were successfully amplified and analyzed from a sample of 60 DNA samples extracted from lesions originating from Cameroon. For each locus, the same allele sizes and a similar number of null alleles (range from 0 to 5% (3/60) across loci) were detected in comparison with the Cameroonian sample of cloned isolates. However, multiple allele patterns (generally with two alleles) for several loci were detected for 13.3% (8/60) of the DNA samples. Since *M. fijiensis* is haploid, these patterns certainly came from the co-alescence of lesions initiated by different spores. Such samples must be considered as missing data since multiloci genotypes could not be determined. However, this drawback could be minimized through selection of lesions sufficiently isolated from each other on leaves, and the rapid, low-cost extraction method presented here will easily allow increased sample sizes to compensate for missing data.

CONCLUSIONS

The combination of the two technical developments presented in this note allows us to optimize our genotyping capacities in *M. fijiensis* for analysis of large samples. The tetranucleotide microsatellites developed are technically easy to score from local to global geographic scales, and the possibility of using a common set of microsatellites at different locations opens up new prospects of analysis to understand the dispersal processes in *M. fijiensis*. Similar procedures could be employed in other fungal pathogens of plants.

LITERATURE CITED

- ARCHIE, E. A., G. LUIKART, AND V. O. EZENWA. 2009. Infecting epidemiology with genetics: A new frontier in disease ecology. *Trends in Ecology & Evolution* 24: 21–30.
- GOUDET, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *The Journal of Heredity* 86: 485–486.
- HALKETT, F., D. COSTE, G. G. RIVAS PLATERO, M. F. ZAPATER, C. ABADIE, AND J. CARLIER. 2010. Genetic discontinuities and disequilibria in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Molecular Ecology* 19: 3909–3923.
- JONES, D. R. 2000. Diseases of banana, abaca and enset. CABI Publishing, Oxon.
- NEU, C., D. KAEMMER, G. KAHL, D. FISHER, AND K. WEISING. 1999. Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. *Molecular Ecology* 8: 523–525.
- RAYMOND, M., AND F. ROUSSET. 1995. An exact test for population differentiation. *Evolution; International Journal of Organic Evolution* 49: 1280–1283.
- ROZEN, S., AND H. J. SKALETSKY. 2000. Primer3 on the WWW for general users and for biologist programmers. In S. Krawetz and S. Misener [eds.], *Bioinformatics methods and protocols: Methods in molecular biology*, 365–386. Humana Press, Totowa, New Jersey.
- STOREY, J. D. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 64: 479–498.
- YANG, B. J., AND S. B. ZHONG. 2008. Fourteen polymorphic microsatellite markers for the fungal banana pathogen *Mycosphaerella fijiensis*. *Molecular Ecology Resources* 8: 910–912.
- ZAPATER, M. F., M. DUCHEMIN, J. F. DUSSART, D. COSTE, P. BROTTIER, AND J. CARLIER. 2008. Microsatellite markers for the fungal banana pathogens *Mycosphaerella fijiensis*, *Mycosphaerella musicola* and *Mycosphaerella eumusae*. *Molecular Ecology Resources* 8: 1121–1125.

Résumé

La dispersion est un processus clef dans la dynamique et l'évolution des populations naturelles. En plus de son rôle primordial dans les processus de colonisation, la dispersion influence également les processus d'adaptation des organismes. Chez les pathogènes, une meilleure compréhension des processus de dispersion apparaît de ce fait être un enjeu majeur pour mieux les contrôler. Durant cette thèse, nous avons étudié les processus de dispersion et quantifié les flux de gènes qui en découlent chez le champignon parasite du bananier *Mycosphaerella fijiensis*. Cette étude a été réalisée à l'échelle locale d'un bassin de production du Cameroun (la région dite du Moungo) et nous avons combiné plusieurs approches complémentaires considérant différentes échelles spatio-temporelles. Dans un premier temps, nous avons décrit, à l'aide de marqueurs génétiques neutres, la structuration spatiale des populations de *M. fijiensis* dans la région du Moungo qui présente différentes barrières potentielles à la dispersion. Nous n'avons décelé aucun effet du paysage ni de la distance géographique sur la structuration génétique. Cependant, une rupture spatiale dans les fréquences alléliques, vraisemblablement de nature historique a été mise en évidence. Ces résultats suggèrent l'existence de grandes populations de *M. fijiensis* s'écartant de l'équilibre mutation-dérive. Dans un second temps, nous avons utilisé la théorie des clines génétiques pour étudier les forces à l'origine de la mise en place et de l'évolution de gradients spatiaux de fréquences alléliques. En particulier, l'analyse de la variation spatio-temporelle de la discontinuité génétique précédemment détectée par un modèle de clines neutres nous a permis d'estimer l'intensité des flux géniques ($\sigma = 1175$ m/génération). Finalement, nous avons mesuré la distribution des distances de dispersion des deux types de spores produites par *M. fijiensis* à partir d'une source d'inoculum primaire. Cette expérimentation nous a permis de confirmer que les ascospores participent à une dispersion à grande distance alors que les conidies sont impliquées dans une dispersion à très courte distance. Nous avons estimé une distance moyenne de dispersion de 3,12 et de 283 mètres/génération respectivement pour les conidies et les ascospores et montré que le noyau de dispersion des ascospores est caractérisé par une queue lourde. Cette thèse a permis de préciser comment *M. fijiensis* se disperse et les estimations réalisées pourront être intégrées dans des modèles théoriques afin de mieux comprendre l'évolution des résistances aux fongicides et de définir des stratégies durables d'utilisation raisonnée des traitements chimiques.

Mots clefs

Dispersion, colonisation, flux de gènes, structure génétique spatiale, clines, noyau de dispersion, résistance aux fongicides, *Mycosphaerella fijiensis*, bananier.

Summary

Dispersal is a key process for both the dynamics and evolution of natural populations. In addition to being crucial for colonization, dispersal also influences the processes occurring during adaptation. For pathogens, a better understanding of dispersal processes may improve our capacity to control the diseases that they cause. In this thesis, we studied dispersal processes and quantified gene flow in the banana plant pathogen *Mycosphaerella fijiensis* at the local scale of a production area in South-West Cameroon (named Moungo). For this purpose, several approaches differing in the spatio-temporal scale to which they refer were combined. First, neutral markers were used to describe the spatial genetic structure of this pathogen in the Moungo area, which includes several potential ecological barriers to dispersal. No effects on genetic structure of landscape elements or geographical distance were found. However, we detected a spatial break in allelic frequencies that appeared to be explained by an historical event. This result suggests the existence of large *M. fijiensis* populations out of the mutation-migration-drift genetic equilibrium. Second, genetic cline theory was applied to study the evolutionary forces implicated in the installation and evolution of spatial gradients in allelic frequencies. More specifically, we analysed the spatio-temporal variation of the genetic discontinuity previously detected through a neutral cline model to estimate the intensity of gene flow in this area ($\sigma = 1175$ m/generation). Lastly, we measured the distribution of dispersal distances of *M. fijiensis* spores from a primary source of inoculum was. Such an experiment allowed us to confirm that conidia are implicated in short-distance dispersal whereas ascospores are responsible for spread of the disease over longer distances. The estimated mean dispersal distance travelled by spores was 3.12 and 283 metres/generation for conidia and ascospores, respectively, and the ascospore dispersal kernel was shown to be fat-tailed. This thesis adds to global knowledge of *M. fijiensis* dispersal and the measures of dispersal estimated in this work will be useful in parameterizing models aimed at a better understanding of the spatial patterns of fungicide resistance evolution under different management strategies.

Key words

Dispersal, colonization, gene flow, spatial genetic structure, clines, kernel of dispersal, fungicide resistance, *Mycosphaerella fijiensis*, banana plant.